



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number 97477

TO: William W Moore
Location: CM1/10A09/10D01
Art Unit: 1652
Tuesday, July 01, 2003

Case Serial Number: 014896

From: Toby Port
Location: Biotech-Chem Library
CM1-6A04
Phone: 308-3534

toby.port@uspto.gov

Search Notes

Dear Examiner Moore,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



97477

STIC-Biotech/ChemLib

From: Moore, William
Sent: Wednesday, June 25, 2003 3:31 PM
To: STIC-Biotech/ChemLib
Subject: Sequence Search request for 10/014,896

OK

Please search SEQ ID NO:1 in 10/014,896 in the following ways:

1. N2N standard, Search SEQ ID NO:1 in the EST, GenEmbl, N-GeneSeq, US patented, and USPGpublished nucleic acid sequences.
2. N2N interference, Search SEQ ID NO:1 in the US pending main and pending new application nucleic acid sequences.
3. N2P standard, Search SEQ ID NO:1 **against** the PIR, SwissProt, SPTREMBL, A-GeneSeq, US patented, and USPGpublished amino acid sequences.
4. N2P interference, Search SEQ ID NO:1 **against** the US pending main and pending new application amino acid sequences.

Thanks,
William W. Moore, Art Unit 1652
Telephone: 703.308.0583
Office: 10A09 Crystal Mall 1
Mailbox: 10D01 Crystal Mall 1
william.moore@uspto.gov

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

```

RESULT 1
US-10-014-896-2
; Sequence 2, Application US/10014896
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchan
; APPLICANT: Miranda, Maricar
; APPLICANT: Fiddle, Carl Johan
; TITLE OF INVENTION: Novel Human Proteases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0280-USA
; CURRENT APPLICATION NUMBER: US/10/014,896
; CURRENT FILING DATE: 2001-12-11
; PRIORITY APPLICATION NUMBER: US 60/255,567
; PRIORITY FILING DATE: 2000-12-11
; NUMBER OF SEQ ID NOS: 4

```


OY 859 GGAAGGGGACAGTGTGCTATTCAGCAACTGGCAATAG 903
DB 202 GlysSerGlyThrValThrValLeuGlnGlnLeuAlaAsnGlu 216

RESULT 5

US-09-902-540-10715
Sequence 10715, Application US/09902540
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Miesgen, Roger C.
TITLE OR INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIORITY FILING DATE: 2000-07-10
PRIORITY FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 10715
LENGTH: 408
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-09-902-540-10715

Alignment Scores:
Pred. No.: 1,47e-65 Length: 408
Score: 796.50 Matches: 170
Percent Similarity: 60.93% Conservative: 78
Best Local Similarity: 41.77% Mismatches: 147
Query Match: 28.84% Indels: 12
DB: 23 Gaps: 7

US-10-014-896-1 (1-1509) x US-09-902-540-10715 (1-408)

OY 264 TACAGTGTGACGACCCAGCTTTATCCAGCAAGTGTGCAAGAGTATATCCAGCTGT 323
DB 5 TyrProArgLeuHisGlnAla-LeuLysArgGluProValGlyAlaHisSerValLeu 24
OY 324 CACATCCAGGCTGGAGCCAGCCAGCTTCAGCAAGCTTCAGCTTCAGCTTCAGCTTC 383
DB 24 ThrTrpThrGlyThrAlaSerLeuArgProAlaLeuLeuGlyHisLeuAspVal 44
OY 384 GGTG-----CCTGCCCTGAGAGAGGCTGGAGAGTCCCATCTCTGGGTGGA 434
DB 44 ValProValGluProGlyThrGluAlaSerTrpThrHisProProGlySerGlyLeuVal 64
OY 435 GCGTATGCGCTCATCTATGTGTCGGGGACACCTGAGAGCAAGAACTCTGATGCGAT 494
DB 64 AlaAspGlyTyrValTyrGlyArgGlyAlaLeuAspArgLysGlySerValPheGlyTyr 84
OY 495 ACTGAGGCGCTGAGTCTGTCGTCGATGAGAGTACATCCCGGAGATCTTCTTCAT 554
DB 84 LeuGlnSerValGluAlaLeuLeuAlaGlyPheGlnProLysArgThrValLeuLeu 104
OY 555 TCTCTGGCCATGATGAGAGCTCATCGG---ACAGGGGCTCAGAGATCAGCCCT 611
DB 104 ValPheGlyGlyAspArgGluGluValGlyArgGluValGluAlaGluAlaMetAlaLysLeu 124
OY 612 GCTACAGTCAAGGGGCTGACAGCTAGCTTCATTTGTGAGAGGGGGCTTCATTTGGA 671
DB 124 LeuValArgGluArgGlyValThrLeuGlnSerValLeuAspArgGlyMetLeuVal 144
OY 672 TGAATTCATCTCACTCAAGAGCCATGCTTCAGCTTCAGAGAGGGTTC 731
DB 144 rGlyThrValProGlyValAlaSerProValAlaLeuValGlyValSerGlyLysGlyPhe 164
OY 732 CATGAACCTGATGCTGCAAGTAAAGATGATCTCAGGACCACTTGTAGCTCCCAAGAG 791
DB 164 eAlaSerAlaGluLeuValAlaAspGlyGluGlyHisSerMetProProProGlu 184
OY 792 GACACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851
DB 184 nThrAlaValGlyValLeuSerArgAlaLeuSerArgLeuGlnAspAlaProMetProAl 204
OY 852 CATATTTGGAAGCGGACAGTGTGCTATTCAGCAACTGGCAAT--GAGTTTCC 908
DB 204 aLysLeuArgGlyGlySer-----ArgAlaLeuPheGlnPheAlaGlyProGluMetGly 222
OY 909 CTTCCCTGTCATATATATCTGAGCAAGCCAGTATTTGAGCACTTATAGCAAGTT 968
DB 222 yPheGlyMetArgThrLeuPheAlaAsnLeuTrpLeuPheGlnProLeuValLeuArgGly 242
OY 969 TATGAGAGAAATCCCTTAAATGCAATATATGAGAGAGCCAGGACATCCATAT 1028
DB 242 nLeuThrAlaLysAlaThrThrAsnAlaAlaValArgThrThrThrAlaAlaThrMetPhe 262
OY 1029 CAAAGCGGGGTCAAGTTCATGTCATCCCGCCAGAGCCAGGACGACATCACTTCG 1088
DB 262 eGluGlySerGlnArgAspAsnValLeuProAlaArgAlaValAlaAsnPheArg 282
OY 1089 GATTCACCTGGACAGACAGTCCCAAGAGTCCCTAGAACTCAGAGAAATGTTGGCTGA 1148
DB 282 gLileuProGlyAspSerValAlaGlyValLeuGlnLysValArgValValAspAs 302
OY 1149 TACAGAGTCCAGTTCATGTTGAGTGCCTTT--GACCCCTCCCGTCAGCCCTTC 1205
DB 302 ProArgValLysValArgThrLeuGlnPheLeuSerGlnProSerProValSerArgLe 322
OY 1206 TGATGACAGGCTTGGGCTACAGCTGTCGCCAGACCTGACATCCGCTTCCCGGA 1265
DB 322 tAspSerGlnAla-----TrpSerGlnLeuGlnArgSerValArgGlyValPheProAs 340
OY 1266 AGTCAAATTTCTGCCCGCAGTACTTCTATGGCAACAGACAGAGGATCTTCAAA 1325
DB 340 pVal---ValAlaProGlyTrpLeuMetLeuGlyAlaThrAspSerArgTrpPheThr 359
OY 1326 CCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1385
DB 359 yLeuSerGlnAsnValTyrArgPheMetProLeuArgLeuAspGlyAlaAspLeuSerArg 379
OY 1386 CATCCATGAGTCAACGAGAAATCTGACGTCGCAAGCTTATGAGAGCCAGTGAATTT 1445
DB 379 gLeuHisGlyLysAspArgGluValSerValLysGlyTyrAlaAspAlaValArgPheTyr 399
OY 1446 CTTTGAGTTGATTCAGAT 1464
DB 399 rAlaGlnTyrValArgAsn 405

RESULT 6

PCT-US02-09921-1008
Sequence 1008, Application PC/TUS0209921
GENERAL INFORMATION:
APPLICANT: INCITE GENOMICS, INC.
APPLICANT: DAFFO, Abel
APPLICANT: JONES, Anissa L.
APPLICANT: TRAN, Anna-Phung B.
APPLICANT: DAHL, Christopher R.
APPLICANT: GIETZEN, Daryl
APPLICANT: CHINN, Joyce
APPLICANT: DUFOUR, Gerard E.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: YU, Jimmy Y.
APPLICANT: TUASON, Olivia
APPLICANT: YAP, Pierre E.
APPLICANT: AMSHEY, Stefan R.
APPLICANT: DAUGHERTY, Sean C.
APPLICANT: DAM, Tam C.
APPLICANT: LIU, Tommy F.
APPLICANT: NGUYEN, Duy-Viet An
APPLICANT: KLEBEELD, Yael
APPLICANT: GERSTIN JR., Edward H.
APPLICANT: PERALTA, Careyana H.
APPLICANT: DAVID, Marie H.
APPLICANT: LEWIS, Samantha A.
APPLICANT: CHEN, Alice J.

```

1  APPLICANT: PANZER, Scott R.
2  APPLICANT: APPLICANT: HARRIS, Bernard
3  APPLICANT: APPLICANT: FLORES, Vincent
4  APPLICANT: APPLICANT: MARWAHA, Rakesh
5  APPLICANT: APPLICANT: LO, Audrey
6  APPLICANT: APPLICANT: LAN, Ruth Y..
7  APPLICANT: APPLICANT: DRASHKA, Michael
8  TITLE OF INVENTION: SECRETORY MOLECULES
9  FILE REFERENCE: PT-1232 PCT
10 CURRENT APPLICATION NUMBER: PCT/US02/09921
11 CURRENT FILING DATE: 2002-03-27
12 PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
13 60/291,829; 60/299,428; 60/300,001; 60/299,776
14 PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
15 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
16 NUMBER OF SEQ. ID NOS: 1146
17 SOFTWARE: PERL Program
18 SEQ ID NO 1008
19 LENGTH: 154
20 TYPE: PRT
21 ORGANISM: Homo sapiens
22 FEATURE:
23 NAME/KEY: misc.feature
24 OTHER INFORMATION: Incyte ID No.: LG:369861.5.off1.2001JUN22
25 PCT-US02-09921-1008

```

Db 98 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTyrPgluValProPro 117
QY 421 TTCTCTGGG 429
Db 118 PheSerGly 120

RESULT 8
US-60-311-261-2481
; Sequence 2481, Application US/60311261
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhwei
; APPLICANT: Meng, Gezh
; APPLICANT: Ma, Yungling
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
; FILE REFERENCE: 806
; CURRENT APPLICATION NUMBER: US/60/311, 261
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: pC_FL_genes Version 5.0
; SEQ ID NO 2481
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-311-261-2481

Alignment Scores:
Pred. No.: 1.02e-43 Length: 120
Score: 562.50 Matches: 114
Percent Similarity: 79.72% Conservative: 0
Best Local Similarity: 79.72% Mismatches: 0
Query Match: 20.37% Indels: 29
DB: 27 Gaps: 1

US-10-014-896-1 (1-1509) x US-60-311-261-2481 (1-120)

QY 1 ATGGCTCAGCGGTGCTTGGTGGCCCTGGGTGCTAGTTTTCCT 60
Db 7 MetAlaGlnArgCysValAlaCysValLeuAlaLeuValAlaMetLeuLeuValAlaPhePro 26

QY 61 ACCGCTCCAGATCATGGGCGCCGAGGAGCGGGGAGCATCAAGGCGCTCGCGAATCCCT 120
Db 27 ThrValSerArgSerMetCglYrProArgSerGlyGlnHisGlnArgAlaSerArgIlePro 46

QY 121 TCTCAGTTCAGCAAGAGGAGACCGCTCGGATGAAAGCGCGTGAAGGTCCTACATCCAG 180
Db 47 SerGlnPheSerLysGluGluValArgValAlaMetLysGlnAlaLeu----- 61

QY 181 ATTCACAGGTGACTTTTGTGCTGTGAGAGTCCATFACACAGCCCTGGCTGATTCGGA 240
Db 61 ----- 61

QY 241 AATAACATTCATAAAGTCTTCTTACAGTGTGTACAGCAGCTTATCCAGCATGATGC 300
Db 62 -----LysValPheProThrValValSerThrPheIleGlnHisGluVal 77

QY 301 GTGGAGAGTATAGCCACCTGTTCATATTCACAGCTGGAGCCACCTTGACGCCCTAC 360
Db 78 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr 97

QY 361 CTGCTGATGCTCAGTCTTATGATGGTGGCTGCGCCCTGGAAGAGCTGGAGGTCGCCCA 420
Db 98 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTyrPgluValProPro 117

QY 421 TTCTCTGGG 429
Db 118 PheSerGly 120

RESULT 9
US-10-179-131-8920
; Sequence 8920, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORR
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 8920
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-8920

Alignment Scores:
Pred. No.: 9.31e-41 Length: 581
Score: 535.00 Matches: 164
Percent Similarity: 46.65% Conservative: 80
Best Local Similarity: 31.36% Mismatches: 197
Query Match: 19.37% Indels: 82
DB: 25 Gaps: 20

US-10-014-896-1 (1-1509) x US-10-179-131-8920 (1-581)

QY 118 CTTTCTCAGTTCACCAAGAGGAGCGGTGCGCATG----- 153
Db 64 ProGluSerTyrTyrLysAspAsnSerThrValLeuGluIleLeuHisAspLysTyr 83

QY 154 ---AAAGAGCGCGTGAAG-----GCTGCCATTCAGATTCACACAGTTCATTCAGC 201
Db 84 LysLysGluSerIleLysArgLeuAlaGlyValIleGlnValAspThrGlnValAlaPheAsp 103

QY 202 TCTGAGAGTCCATTCATACAGCC-----CTGGCTGAGTTCGGAAAA 263
Db 104 LysGlnProAlaValAlaAspAlaProGluValTrrAlaLysPheAlaLysPheHisAsp 123

QY 244 TACATTCATAAAGTCTTCTTCTACAGTGTGACAGCAGCTTATCCAGATGAATCGTG 303
Db 124 TyrLeuGluGlnThrPheProLeuValTyrLysAsn-----LeuLysValThrLysVal 161


```
QY 475 AAGAGCTGTGATGACATTACAGAGGCTTCGAGCTCTGATGACGAAGTACATC 534
    ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 237 LysAsnValIleuIleAlaIleuGluThrIleuGluIleuAlaLysGlyTyrGln 256
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 535 CCCGAGAGATCTTCTTCATTTCTCTGGCCATGATAGAGTACATCAGGG--ACAGGG 591
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 257 ProLysArgSerIleuAlaIleuAlaPheGlyPheAspGluIleuThrSerGlyTyrHisGly 276
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 592 GCTCAGAGATCTCAGCCCTGCTACAGTCAAGGGGGCTCAG-----CTAGCTTCAT 645
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 277 AlaAlaHisIleuGlyTyrGluGluGluThrPheGlyLysAspSerValTyrAlaLeu 296
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 646 GTGAGAGGGGGCTTCATCTGATGATTTTCATCTCCATCTCAAGAACCCCATCGCC 705
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 297 IleAspGluIleuAlaGlyLeuThrValGlnGluIleuThrAsn-----ThrIleValAla 314
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 706 TTGATTCAGCTTCAGAGAGGGTTCATGACCTCAGCTCAGTCAAGTAAACATGACTCA 765
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 315 LeuProGluThrAlaGluLysGlyTyrValAspIleGluValGluIleuThrProGly 334
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 766 GGCACACTTCAGCTCTCCAAAGAGACAAAGATGGCATCTGACAGCTGCTCAGC 825
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 335 GlyHisSerSerIleProProAspProThrSerIleGlyIleIleSerGluLeuGlyTyr 354
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 826 CGATTTGAGAGACACCA---ATGCTATCATATTGGAAGCGGAGACGTGTGACTGTA 882
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 355 IleIleGluLysAspProGlySerProLeuLeuProGluAsnProIleLeuAsnPhe 374
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 883 TTGCAGCACTGSCA-----AATGATGTTCC-----TTCCCTGTCAT 921
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 375 AlaGlnCysLeuAlaIleuHisAspProLysAsnHisIleProSerPheLysAla 394
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 922 AATAATC-----CTGAGCAACCCATGCTATTTGAACCACTAATAAGC 963
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 395 IleuAlaGluAlaGlyTyrAspLysPheAlaAsnSerLysLeuValGluGlyIle----- 412
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 964 AGGTTTATGAGAGAAATCCCTTAACCAATGCAATAATCAGAGCAACCCAGGCACTAC 1023
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 413 -----SerLysAsnArgLeuThrLysTyrLeuIleArgThrSerGlnAlaLeuAsp 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1024 AATATTCAGAGAGGGTCAATGTAATGTCATCCCCCAAGTGGCCGAGCCAGCACTGCAAC 1083
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 430 IleIleAsnGlyGluLysAlaAsnAlaLeuProGluHisValLysLeuLeuValAsn 449
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1084 TTCCGATTCACCTTGAGACAGACAGTCCAAAGAGTCTCTAGA---CTCAGCAAGAACAT 1140
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 450 HisArgValAlaIleGlyThrSerValAlaGluValGlnGluHisPheValSerArgVal 469
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1141 GTGGCTGATTAACAGAGTCCAGTCCAGTGTGTGAGTGGCTTT----- 1182
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 470 ValGluValAlaLysArgHisGlyLysLeuSerValSerAlaPheGlyLysAspValLeuLys 489
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1183 -----GACCCCTC 1191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 490 ValLysAsnAspSerGlyLeuPheAsnValThrAsnPheAlaGlyPheLeuAsnAlaAla 509
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1192 CCCGTCAGCCCTTGATGACAGAGCTTGGGCTAC-----CAGCTGCTCCGCGAGACC 1245
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 510 ProValIleThrProHisAsnAspThrValIleProGluTyrLeuSerGlyValThrLysHisVal 529
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1246 GTACAGTCC--GTCTCCCGGAGTCAAT-----ATTACTGCCCCAGTTACTTCT 1293
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 530 TyrGluAspLeuValPheProGluIleLysTyrProValIleThrAlaProAlaIleMet 549
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1294 ATTTGCAACAGAGAGCGATCTTTTAAACACACAGCTGAGATGATGATGATGATGATGAT 1353
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 550 ThrGlyAsnThrAspThrAlaGlnHisTyrThrAsnLeuThrAlaGlnIlePheArgPheThr 569
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1354 CCCCATCTACATGACAGCTGAGACATCTC-----AAAGCATCATGAGATCAAGAG 1404
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 570 ProAlaPheIle-----GlyAspPheIleGlyLysThrHisIleHisSerValAspGlu 587
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
QY 1405 AAAATCTCAGTCCAGCCCTATGAGACCCCAAGTGAATTCATCTTGAATTCAGAT 1464
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 588 LysLeuProPheAspSerHisIleuGlnIleuGlnAlaIleuThrPheTyrGluTyrIleGlnAla 607
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1465 GCTGACACA 1473
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 608 IleAspSer 610
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 11
US-60-096-409-17521
; Sequence 17521, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17521
; LENGTH: 616
; TYPE: PRN
; ORGANISM: Candida albicans
; US-60-096-409-17521

Alignment Scores:
Pred. No.: 1,85e-40 Length: 616
Score: 532.00 Matches: 163
Percent Similarity: 46.65% Conservative: 81
Best Local Similarity: 31.17% Mismatches: 197
Query Match: 19,26% Indels: 82
DB: 27 Gaps: 20

US-10-014-896-1 (1-1509) x US-60-096-409-17521 (1-616)
QY 118 CTTTCAGTTCAGCAAGAGGAAAGCGTCCGATG----- 153
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 99 ProGluSerTyrTyrLysAspAsnSerThrValIleGluIleuHisAspLysTyr 118
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 154 ---AAAGAGCGCTGAAA-----GCTGCATCCAGATTCCAAAGTCACTTTTAC 201
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 119 LysLysGluSerIleLysArgLysAlaIleGlnValAspThrGlnValPheAsp 138
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 202 TCTGAGAGTCCAACTACTACAGCC-----CTGGCTAGTTCGGAAAA 243
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 139 LysGlnProAlaValAspAspAlaProGlnValTyrAlaLysPheAlaLysPheHisAsp 158
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 244 TACATTCATAAGTCTTCTCTACAGTGCAGACAGCTTATCCAGATGAATGCTG 303
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 159 TyrLeuGluIleuThrPheProLeuValTyrLysAsn-----LeuLysValThrLysVal 176
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 304 GAAGAGTAAAGCCACTGTTCACTATTCACAAAGCTCGAGCCCAAGTTCGACCCCTACCTG 363
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 177 AsnThrTyrGlyLeuValTyrHisTyrLysGlySerAspLysSerLeuLysProValLeu 196
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 364 CTGAGTCTCACTTGAATGTGTGTCCTGCTCCGCTCCGCTCGAAGAA-----GGCTGGAGAGTG 414
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 197 LeuThrAlaHisGlnAsnThrValProValIleLysAspThrLeuLysAspThrThrTyr 216
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 415 CCCCATCTCTCGGCTTGAGAGCGGATGAGCTCATGATGATGATGATGATGATGATGATGAT 474
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 217 ProThrGluGluGlyHisTyrAspGlyGluTyrIleTyrGlyLysGlyAlaAlaAspCys 236
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 475 AAGAGCTGTGATGACATTACAGAGGCTTCGAGCTCTGATGACGAAGTACATC 534
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 237 LysAsnValIleuIleAlaIleuGluThrIleuGluIleuAlaLysGlyTyrGln 256
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 535 CCCGAGAGATCTTCTTCATTTCTCTGGCCATGATAGAGTACATCAGGG--ACAGGG 591
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 257 ProLysArgSerIleuAlaIleuAlaPheGlyPheAspGluIleuThrSerGlyTyrHisGly 276
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 592 GCTCAGAGATCTCAGCCCTGCTACAGTCAAGGGGGCTCAG-----CTAGCTTCAT 645
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 277 AlaAlaHisIleuGlyTyrGluGluGluThrPheGlyLysAspSerValTyrAlaLeu 296
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 646 GTGAGAGGGGGCTTCATCTGATGATTTTCATCTCCATCTCAAGAACCCCATCGCC 705
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 297 IleAspGluIleuAlaGlyLeuThrValGlnGluIleuThrAsn-----ThrIleValAla 314
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 706 TTGATTCAGCTTCAGAGAGGGTTCATGACCTCAGCTCAGTCAAGTAAACATGACTCA 765
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 315 LeuProGluThrAlaGluLysGlyTyrValAspIleGluValGluIleuThrProGly 334
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 766 GGCACACTTCAGCTCTCCAAAGAGACAAAGATGGCATCTGACAGCTGCTCAGC 825
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 335 GlyHisSerSerIleProProAspProThrSerIleGlyIleIleSerGluLeuGlyTyr 354
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 826 CGATTTGAGAGACACCA---ATGCTATCATATTGGAAGCGGAGACGTGTGACTGTA 882
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 355 IleIleGluLysAspProGlySerProLeuLeuProGluAsnProIleLeuAsnPhe 374
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 883 TTGCAGCACTGSCA-----AATGATGTTCC-----TTCCCTGTCAT 921
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 375 AlaGlnCysLeuAlaIleuHisAspProLysAsnHisIleProSerPheLysAla 394
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 922 AATAATC-----CTGAGCAACCCATGCTATTTGAACCACTAATAAGC 963
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 395 IleuAlaGluAlaGlyTyrAspLysPheAlaAsnSerLysLeuValGluGlyIle----- 412
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 964 AGGTTTATGAGAGAAATCCCTTAACCAATGCAATAATCAGAGCAACCCAGGCACTAC 1023
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 413 -----SerLysAsnArgLeuThrLysTyrLeuIleArgThrSerGlnAlaLeuAsp 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1024 AATATTCAGAGAGGGTCAATGTAATGTCATCCCCCAAGTGGCCGAGCCAGCACTGCAAC 1083
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 430 IleIleAsnGlyGluLysAlaAsnAlaLeuProGluHisValLysLeuLeuValAsn 449
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1084 TTCCGATTCACCTTGAGACAGACAGTCCAAAGAGTCTCTAGA---CTCAGCAAGAACAT 1140
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 450 HisArgValAlaIleGlyThrSerValAlaGluValGlnGluHisPheValSerArgVal 469
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1141 GTGGCTGATTAACAGAGTCCAGTCCAGTGTGTGAGTGGCTTT----- 1182
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 470 ValGluValAlaLysArgHisGlyLysLeuSerValSerAlaPheGlyLysAspValLeuLys 489
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1183 -----GACCCCTC 1191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 490 ValLysAsnAspSerGlyLeuPheAsnValThrAsnPheAlaGlyPheLeuAsnAlaAla 509
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1192 CCCGTCAGCCCTTGATGACAGAGCTTGGGCTAC-----CAGCTGCTCCGCGAGACC 1245
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 510 ProValIleThrProHisAsnAspThrValIleProGluTyrLeuSerGlyValThrLysHisVal 529
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1246 GTACAGTCC--GTCTCCCGGAGTCAAT-----ATTACTGCCCCAGTTACTTCT 1293
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 530 TyrGluAspLeuValPheProGluIleLysTyrProValIleThrAlaProAlaIleMet 549
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1294 ATTTGCAACAGAGAGCGATCTTTTAAACACACAGCTGAGATGATGATGATGATGATGATGAT 1353
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 550 ThrGlyAsnThrAspThrAlaGlnHisTyrThrAsnLeuThrAlaGlnIlePheArgPheThr 569
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1354 CCCCATCTACATGACAGCTGAGACATCTC-----AAAGCATCATGAGATCAAGAG 1404
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 570 ProAlaPheIle-----GlyAspPheIleGlyLysThrHisIleHisSerValAspGlu 587
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

Db      277 AlaAlaHisLIleGlyLysTyrLeuGlnGluThrPheGlyGlnAspSerValTyrAlaLeu 296
Oy      646 GTGGAGACAGGGGGGCTCATCTTGGATGATTTATCTCTAACTTCAAGAGCCATGGCC 705
Db      297 TLeaRgGlnGlyAlaGlyLeuThrValGlnGluLeuThrAsn-----ThrLeuAla 314
Oy      706 TTGATTTGACAGTCTCAGAGAAAGGGTTCATGACAGCCATGCTGCAAGTAAACATGACTTCA 765
Db      315 LeuProGluThrAlaGluLysGlyTyrValAspRIleGlnValaGluLeuThrThrProGly 334
Oy      766 GGCCACTTCTACGCTCTCCCAAGAGACAAATGGAGATCCCTGGACCTGGACCTGGTCAAGC 825
Db      335 GlnHisSerSerIleProProAspProThrSerIleGlyYIleIleSerGluLeuGlyTyr 354
Oy      826 CGATGGAGAGACAGACCA-----ATGCGTATGATATTGGAAAGGGGACAGTGGTAGCTGA 882
Db      355 TleIleGluLysAspProTyrSerProLeuLeuProGluLysProIleLeuLysPhe 374
Oy      883 TTGCAGCACTGGCA-----AATGAGTTTCCC-----TTCCCTGTCTAAT 921
Db      375 AlaGlnCysLeuAlaLeuHisAspProLysAlaAsnIleLeuSerSerPheLeuAla 394
Oy      922 ATTAAC-----CTAGACAAACCATGGCTATTGGACCACTTTTAAC 963
Db      395 TleLeuAlaGlyAlaGlyTyrAspLysPheAlaAsnSerLysLeuAlaGluGlyLe----- 412
Oy      964 AGGTTATTTGAGAGAAATCCCTTACACAAATGACAAATATGACAGCAACAGCCACATCCAC 1023
Db      413 -----SerLysAsnAlaGluThrLysTyrLeuIleGlnThrSerGlnAlaLeuAsp 429
Oy      1024 ATATTTCAAAGCAGGGGGTAACTTAAATGTCATCCGCCCACTGGCCAGGCCCACTGAAC 1083
Db      430 TleIleAsnGlyLysLysAlaAlaAsnAlaLeuProGluHisValLysLeuLeuValAsn 449
Oy      1084 TTCCGGATTACCCCTGGACAGACAGACAGTCCAAAGGCTCCAGAA-----CTACAGAAAGCAAT 1140
Db      450 HisArgValAlaIleGlyThrSerValAlaGluLysGlnHisPheAlaSerArgVal 469
Oy      1141 GTGGCTGATTAACAGAGTCCAGTCCAAATGCTTGAGTCCCTT----- 1182
Db      470 ValGluValAlaLysArgHisGlyLeuSerValSerIleAspPheGlyLysAspValLeuLys 489
Oy      1183 -----GACCCCTCC 1191
Db      490 ValLysAsnAspSerGlyLeuPheAsnValThrAsnPheAlaGlyPheLeuAsnAlaAla 509
Oy      1192 CCGCTACAGCCCTTGTGATGACAGAGCCCTGGGCTAC-----CAGCTGCTCCGCAAGAC 1245
Db      510 ProValThrProThrAsnAspThrValTyrProLysLeuSerGlyLysThrLysAlaVal 529
Oy      1246 GTACAGATCC-----GTCTTCGGGAGAGTAAAT-----ATTATCCGCCCACTTACTCT 1293
Db      530 TyrGlnAspLeuValPheProGluLysGlyTyrProValIleThrAlaProAlaIleMet 549
Oy      1294 ATTGGCAACAGACAGAGCCGATTTTTCACAAACCTCAACAGCTGGCATCTTACAGTTTTCAC 1353
Db      550 ThrGlyAsnThrAspThrArgHisTyrTyrAsnLeuThrArgAsnIlePheArgPheThr 569
Oy      1354 CCGATCTAATACAGCCGTGAAGACTC-----AAAGCATTCATGGATGAACAAGAC 1404
Db      570 ProIlePheIle-----GlyAspPheIleGlyLysThrHisIleHisSerValAspCys 587
Oy      1405 AAATGTCAGTCCAAAGCCCTATGAGACCCCAAGTGAATATCATCTTGAATTTGATTTAGAT 1464
Db      588 LysLeuProPheAspSerHisLeuGlnLeuGlnAlaThrPheArgGluTyrIleGlnAla 607
Oy      1465 GCTTGACACA 1473
Db      608 IleAspSer 610

```

```

GENERAL INFORMATION:
APPLICANT: HARE, ROBERTA S.
APPLICANT: SHAW, KAREN J.
APPLICANT: SHIMER JR., GEORGE H.
APPLICANT: KESSLER, MARCO
APPLICANT: NOLLING, JORK
APPLICANT: ZENG, QIANDONG
APPLICANT: GREENE, JONATHAN R.
TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
TITLE OR INVENTION: AND USES THEREFOR
FILE REFERENCE: 2976-4031
CURRENT APPLICATION NUMBER: US/10/179.131
CURRENT FILING DATE: 2002-06-21
NUMBER OF SEQ ID NOS: 10194
SEQ ID NO 7140
LENGTH: 623
TYPE: PRN
ORGANISM: Candida albicans
US-10-179-131-7140

Alignment Scores:
Pred. No.: 7.39e-38 Length: 623
Score: 504.50 Matches: 144
Percent Similarity: 47.728 Conservative: 97
Best Local Similarity: 28.518 Mismatches: 195
Query Match: 18.274 Indels: 69
DB: Gaps: 17

US-10-014-896-1 (1-1509) x US-10-179-131-7140 (1-623)
OY 133 AAGAGAACCGCTGCCATGAAGAAGCCGTGAAAGTGCATTCACAGATTCACAACAGTC 192
Db ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
120 LysSerPheHisThrSerValArgLysLysLeuAlaValLysIleProHidn 139
OY 193 ACTTTAGCTCTGAGAAATCC-----ATACTACAGCCCTGGCTAG----- 234
Db ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
140 IllePheAspGlYmeCisLusnProGlnSerIleValLysSerIleuLysIleuTyrgluleu 159
OY 235 -----TTGGAAAATACATTCAATTAAGACTTTTCCCTACA 267
Db ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
160 AspProArgTrpLysProPheglLysPheHisAspTyrlleugluLysTrpTyroPheLuu 179
OY 268 GTGGTCAGACCAGCTTTATCCAGCATGAGACGCGGAGAGATATACCAACCTGGTCCACH 327
Db ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
180 Val-----HisLysHisLeuIndleuLysLysIleAsnLysPheglYleuValTyrlhr 197
OY 328 ATCCAAGCTCGAGCCCCAGCTTGACAGCCCTACCTGTGATGCTCACTTGAATGGTG 387
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 TrpLysglYlysAspRhrSerLysProIleMetleuTrpAlaHisGlnAspValal 217
OY 388 CCGGCCCTCGAAGA-----GGCTGGAGAGTGGCCCCCATCTCTGGGGTTGAGAGCG 438
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 ProvalProHisgluThrIleAspGlnTrpThrPheProPheProPheglugluGlyrPheAsp 237
OY 439 GATGGCGTCATCTAGTGGGGGAGCACTGACAGCAACAACCTGTGATGAGCATTAAG 498
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
238 GlYserTyrlleuTyrglYargLysAlaSerAspCysLysAsnleuAlaIleAlaMet 257
OY 499 CAGGCGTTGAGCTCTGCTGTGATCAG---AAGTACATCCCCGAGAAATCTTCTTCATT 555
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
258 GlYThrIleIleuLeuLeuclunslupsRhyrPheLysProGlnAlaTyrlleIleLeu 277
OY 556 TCCTGTGGCCATGATAGAGATCATCAGGAGAGGGGCTCAGAGATCTACAGCCCTGCTA 615
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
278 AlaPheglYtyrAspLugluAlaIleaglYlsglYalaIlegluIulleSerAspTyrlleu 297
OY 616 CAGTCAAG-----GGCGTCAAGCTACGCTTCAATGTGACGAGGGG-----GGCTTC 663
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
298 ValAsnLysTyrglYProAspSerIleLeuGlnIleIleAspGlnIleAspGlnIleTytr 317
OY 664 ATCTGTGATCATTCATCTCAATCTCAAGAACGCCATCGGCTTGATGATGAGCTGCAGAG 723
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
318 -----GlnIuIleIleuIleuValLys-----IleuValleuProIleAlaTrpIleGlu 332

```


[illegible]

```

STREET: 501 West Broadway
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Mln95
SOFTWARE: word

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,468
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET.012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Large intestine
FEATURE:
NAME/KEY: s1g_peptide
LOCATION: -23..1
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 9.1
OTHER INFORMATION: seq LVAMLLVFPYVS/RS
US-08-904-468-192

Alignment Scores:
Pred. No.: 3,09e-33 Length: 93
Score: 451.00 Matches: 93
Percent Similarity: 98.94% Conservative: 0
Best Local Similarity: 98.94% Mismatches: 0
Query Match: 16.33% Indels: 1
DB: 13 Gaps: 0

US-10-014-896-1 (1-1509) x US-08-904-468-192 (1-93)

QY 1 ATGGCTCAGCGGTGCGCTGGCTGCTGSCCTGTGGCTAGTACTGCTCTAGTTTCCT 60
Db 1 MetAlaGlnArgCysValCysValLeuAlaLeuValAlaMetLeuLeuValAlaPhePro 20
QY 61 ACCGTCGCCAGATCGATGGGCGCGAGAGCGGGAGCATCAAGAGGCGCTGCGAATCCCT 120
Db 21 ThrValSerTrpSerMetGlyProAlaGlySerGlyGlnHisGlnArgAlaSerArgIlePro 40
QY 121 TCTCAGTTTCAGCAAGAGAACGCGTCCGATGAAGAAGACGCGCTGAAAGGTGCCATCCAG 180
Db 41 SerGlnPheSerTrpGlnGlnArgValAlaMetTrpGlnAlaLeuValGlyAlaIleGln 60
QY 161 ATTCGACAGCGAGCTTTTACCTCTGAGAAAGTCCAACTACACGCGCTGGCTGAGTGGGA 240
Db 61 IleProTrpValTrpPheSerSerGlnTrpSerAsnThrThrAlaLeuAlaGlnPheIle 80
QY 241 AAATACATTCAATAAGCTTTCTCTACACAGTGGTCAGCACCA 280
Db 81 AsnThrPheIleTrpSerPheLeuGlnTrpSerAlaPro 93

RESULT 14
US-09-547-599C-2910
; Sequence 2910, Application US/09547599C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste

```


Tue Jul 1 11:49:11 2003

us-10-014-896-1.n2p.ram

Page 2

[illegible]

OY	1081	AACTTGGGATTCACCCCTGGACAGACAGTCCAGAGAGGTCCTAGACTACAGAAACATT	1140
Dd	332	AspPhaArgIleHisProGlyGlnThrValGlnGluValLeuGluLeuThrLysAsnIle	351
OY	1141	GTGGCTGATTAACAGAGTCCAGTCCAGTGTGTGAGTGCCTTTGACCCCCCTCCGCTACGC	1200
Dd	352	ValAlaAspAsnArgValGlnPheHisValLeuSerLalaPheAspProLeuProValSer	371
OY	1201	CCCTTGAAGAACAGGCGTTGGGCGCTTACCAGCTGCTCGCCGACCCCTACAGTCCGCTTTC	1260
Dd	372	ProSerAspArgLysAlaLeuGluLYTyrGlnLeuLeuArgGlnThrValGlnSerValPhe	391
OY	1261	CCGAGAGCTAAATTAATACGCCGCCAGTTACTCTTAATGGGCAACACAGCCGAGTCTTT	1320
Dd	392	ProGluValAsnIleThrAlaProValThrSerIleLysAsnThrAspSerArgPhePhe	411
OY	1321	ACAACTTACACACTGGCAGTCTTACAGGTTCTACCCCATCTTACATACAGCCTTAAGACTTC	1380
Dd	412	ThrAsnLeuThrThyGlyIleTyrArgPheTyrProIleTyrIleGlnProGluAspPhe	431
OY	1381	AAACGCATCATGAGTGAAGAGAGAAATCTCACTCCAGACCTATGAGACCCCAAGTGAAG	1440
Dd	432	LysArgIleHisGlyValAlaGlnLysArgIleSerValGlnAlaThrGlnIleValLys	451
OY	1441	TTGATCTTGGATCTATATTCAGAAAGCGCAGACAGACAGAGAGGCACTTCTCACTGGAC	1500
Dd	452	PheIlePheGluLeuLeuIleGlnAsnAlaPheThrAspGlnIleProValSerHisLeuHis	471
OY	1501	AAACTG 1506	
Dd	472	LysIleu 473	
RESULT 2			
PCT-US02-32850-22			
Sequence 22, Application PC/TUS0232850			
GENERAL INFORMATION:			
APPLICANT: INCYTE GENOMICS, INC.			
APPLICANT: RAMKUMAR, Jayalakshmi			
APPLICANT: GORVAD, Ann E.			
APPLICANT: BAUGHN, Mariah R.			
APPLICANT: EMERLING, Brooke M.			
APPLICANT: YANG, Junming			
APPLICANT: LEE, Soo Yeun			
APPLICANT: TRAN, Uyen K.			
APPLICANT: BECHTA, Shaanya D.			
APPLICANT: DUGGAN, Brendan M.			
APPLICANT: LEE, Ernestine A.			
APPLICANT: GRIFPIN, Jennifer A.			
APPLICANT: LI, Joana X.			
APPLICANT: SPRAGUE, William W.			
APPLICANT: HAFALIA, April J.A.			
APPLICANT: CHAMULA, Natinder K.			
APPLICANT: LEHR-MASON, Patricia M.			
APPLICANT: KABLE, Amy E.			
APPLICANT: YUE, Henry			
APPLICANT: MARQUIS Joseph P.			
APPLICANT: YAO, Monique G.			
APPLICANT: RICHARDSON, Thomas W.			
APPLICANT: TANG, Y. Tom			
APPLICANT: JIN, Pei			
APPLICANT: CHEN, David			
APPLICANT: BHATTIA, Umesh G.			
APPLICANT: BURRILL, John D.			
APPLICANT: LEE, Sally			
APPLICANT: BLAKE, Julie T.			
APPLICANT: HO, Anne			
APPLICANT: ZHENG, Wenjin			
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE PROTEINS			
FILE REFERENCE: PP-1237 PCT			
CURRENT APPLICATION NUMBER: PCT/US02/32850			
CURRENT FILING DATE: 2002-10-11			
PRIOR APPLICATION NUMBER: US 60/329, 689			
PRIOR FILING DATE: 2001-10-12			

```

1 PRIOR APPLICATION NUMBER: US 60/335,703
2 PRIOR FILING DATE: 2001-10-25
3 PRIOR APPLICATION NUMBER: US 60/348,887
4 PRIOR FILING DATE: 2001-11-09
5 PRIOR APPLICATION NUMBER: US 60/334,145
6 PRIOR FILING DATE: 2001-11-28
7 PRIOR APPLICATION NUMBER: US 60/337,451
8 PRIOR FILING DATE: 2001-12-06
9 PRIOR APPLICATION NUMBER: US 60/340,584
10 PRIOR FILING DATE: 2001-12-14
11 NUMBER OF SEQ ID NOS: 80
12 SOFTWARE: PERL Program
13 SEQ ID NO 22
14 LENGTH: 361
15 TYPE: PRT
16 ORGANISM: Homo sapiens
17 FEATURE:
18 NAME/KEY: misc_feature
19 OTHER INFORMATION: Incyte ID No: 3118830CD1
20 CDT-US02-32850-22

```

```

1 TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
2 FILE REFERENCE: PE-1237 PCT
3 CURRENT APPLICATION NUMBER: PCT/US02/32850A
4 CURRENT FILING DATE: 2002-10-11
5 PRIOR APPLICATION NUMBER: US 60/329,689
6 PRIOR FILING DATE: 2001-10-12
7 PRIOR APPLICATION NUMBER: US 60/335,703
8 PRIOR FILING DATE: 2001-10-25
9 PRIOR APPLICATION NUMBER: US 60/348,887
10 PRIOR FILING DATE: 2001-11-09
11 PRIOR APPLICATION NUMBER: US 60/334,145
12 PRIOR FILING DATE: 2001-11-28
13 PRIOR APPLICATION NUMBER: US 60/337,451
14 PRIOR FILING DATE: 2001-12-06
15 PRIOR APPLICATION NUMBER: US 60/340,584
16 PRIOR FILING DATE: 2001-12-14
17 NUMBER OF SEQ ID NOS: 80
18 SOFTWARE: PERL Program
19 SEQ ID NO 22
20 LENGTH: 361
21 TYPE: PRT
22 ORGANISM: Homo sapiens
23 FEATURE:
24 NAME/KEY: misc feature
25 OTHER INFORMATION: Incyte ID No: 3118830CD1
26 PCT-US02-32850A-22

```


Db 121 ILSErleuGLyHISAspIuGLuSerSerGlyThrGlyAlaGlnArgILeSerAlaLeu 140
QY 613 CTACAGTCAAGGGCGCTCCAGCTAGCCCTTCATGTGGAGCAGAGGGGCGCTTCATCTGGAT 672
Db 141 LeuGlnSerArgILyAlaGlnLeuAlaPheILeValAspGluGlyPheILeuAsp 160
QY 673 GATTTCATTCCTTAACCTCAAGAACCCCATCGCTGANTGTGACGCTCAGAGAGGGTCC 732
Db 161 AspPheILeProAsnPheLysLysProILeAlaLeuThrAlaValSerGluLysGlySer 180
QY 733 ATGACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTTCAGCTCCCAAGAG 792
Db 181 MetAsnLeuMetLeuGlnValAsnMetThrSerGlyHISerSerAlaProProLysGlu 200
QY 793 ACAAGCATTTGGCATTCCTGACGCTGCTGCAGCCGATTTGGACAGACCAATGGCTTAC 852
Db 201 ThrSerILeGlyILeLeuAlaAlaValSerArgLeuGlnGlnThrProMetProILe 220
QY 853 ATATTGGAGCGGGGAGTGTGACTGATTTGACCACTGGCAATGAGTTTCCTTC 912
Db 221 ILephelGlySerGlyThrValValThrValLeuGlnGlnLeuAlaAsnGluPheProPhe 240
QY 913 CCTGTCAATATTAATCTGAGCAACCCATGGCTATTGAAACCACTTAAAGAGTTTATG 972
Db 241 ProValAsnILeILeLeuSerAsnProTrpLeuPheGluProLeuILeSerArgPheMet 260
QY 973 GAGAGAAATCCCTTAACCAATGCATTAATACAGACCAACGCGGCACTTACCATTTCAAA 1032
Db 261 GluArgAsnProLeuThrAsnAlaILeILeArgThrThrAlaLeuThrILePheLys 280
QY 1033 GCAGGGGTCAAGTTCATGTATCCGCCAGTGGGCCAGGCCAGCTCAACTTCGCGATT 1092
Db 281 AlaGlyValILysPheAsnValILeProProValAlaGlnAlaThrValAsnPheArgILe 300
QY 1093 CACCCGTGACAGACATGCTCAAGAGAGCTCTAGAACTCAAGAAACATTTGGCTGATAC 1152
Db 301 HIsProGlyGlnThrValGlnGluValLeuGlnLeuThrLysAsnILeValAlaAspAsn 320
QY 1153 AGAGTCAGTTCATGCTGAGTGCCTTTGACCCCTCCCGCTCAGCCCTTGTGATGAC 1212
Db 321 ArgValGlnPhePheHISValILeSerAlaPheAspProLeuSerProSerAspAsp 340
QY 1213 AAGGCTTGGGCTACCAAGCTCTCCGCCAGACGCTACAGTCCGCTTCCCGGAAGTCAAT 1272
Db 341 LysAlaLeuGlyTYrGlnLeuLeuArgGlnThrValGlnSerValPheProGluValAsn 360
QY 1273 AAT 1275
Db 361 Ile 361
RESULT 7
US-10-218-140-218
; Sequence 218, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/218,140
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curanator Version 1.0
; SEQ ID NO 218
; LENGTH: 166

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (1)...(1)
OTHER INFORMATION: "Xaa" = "any Amino Acid"
US-10-218-140-218
Alignment Scores:
Pred. No.: 3,81e-53 Length: 166
Score: 827.00 Matches: 164
Percent Similarity: 99.39% Conservative: 0
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 29.94% Indels: 0
Gaps: 0
US-10-014-896-1 (1-1509) x US-10-218-140-218 (1-166)
QY 142 CGCGTGGCATGAAGAGCGCTGAAAGTGGCATCCAGATTCGAACAGTACTTTAGC 201
Db 2 ArgValAlaMetLysGlnAlaLeuLysGlyAlaILeGlnILeProThrValThrPheSer 21
QY 202 TCTGAGAGTCCATACTACAGCCCTGGCTGAGTTGCGAAATATCATTAAGCTTT 261
Db 22 SerGluLysSerAsnThrThrAlaLeuAlaGluPheGlyLysTYrILeHISLysValPhe 41
QY 262 CTACAGTGTGCACACACAGCTTATTCAGCATGAAGTGTGAGAGATTAAGCCACTG 321
Db 42 ProThrValValSerThrSerPheILeGlnHISGluValValGlnGluLysSerHISLeu 61
QY 322 TTCACATTCCAAGGCTGGAGCCCGAGCTTGCAGCCCTACCTGTGATGGCTCACTTGAT 381
Db 62 PheThrILeGlnGlySerAspProSerLeuGlnProTYrLeuMetAlaHISPheAsp 81
QY 382 GTGTGTCCCTCCCTGAAGAGAGCTGGAGAGTGGCCCATTCCTGGGTTGAGCGTGAT 441
Db 82 ValValProAlaProGluGlnGlyTrpGluValProProPheSerGlyLeuGluArgAsp 101
QY 442 GCGCTCATCTATGTGCGGGGACACATGAGACACAAGAACTGTGATGGCATACGCGAG 501
Db 102 GlyValILeTYrGlyTrpGlyThrLeuAspLysPheValMetAlaLeuGln 121
QY 502 GCGTGGAGCTCTGCTGATCAGAAATGATCCCGGAAGATCTTTCATTTCTCTG 561
Db 122 AlaLeuGluLeuLeuILeArgLysTYrILeProArgSerPhePheILeSerLeu 141
QY 562 GGCATGATGAGAGTCAACAGGACAGGGGCTCAGAGATCTCAGCCCTGCTACAGTCA 621
Db 142 GlyHISAspGluGluSerSerGlyThrGlyAlaGlnArgILeSerAlaLeuGlnSer 161
QY 622 AGGGCGCTCCAGCTA 636
Db 162 ArgGlyValGlnLeu 166
RESULT 8
US-10-219-382-2481
; Sequence 2481, Application US/10219382
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Meng, Gezhi
; APPLICANT: Ma, Yundong
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
; FILE REFERENCE: 806A
; CURRENT APPLICATION NUMBER: US/10/219,382
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,261
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25

```

:  APPLICANT: Wang, Zhiwei
;  APPLICANT: Meng, Gezh

```

0Y 421 TTCTTGGG 429

Alignment Scores:

Pred. No.: 2,74e-11 Length: -458
 Score: 266.50 Matches: 117
 Percent Similarity: 41.32% Conservative: 71
 Best Local Similarity: 25.71% Mismatches: 212
 Query Match: 9.65% Indels: 55
 DB: 1 Gaps: 18

US-10-014-896-1 (1-1509) x PCT-0502-32727-2506 (1-458)

```

QY 175 ATCCAGATTCACAGCTGATTTTACGCTCTGAGAACTCCAAAT----- 216
DB 35 IIEGIIIEASPSERILASNPHEGLYPROGLINASPLAARGLYGLVALGLMETCYS 54
QY 217 -----ACTACAGCCCTGGCTGAG-----TTCCGAAATATACATTCATAAGCTTT 261
DB 55 HASTYVALTHRCILYLEUENUSPGLIILEGLYVALGLYVALTHRLIENHISGLUSERGL 74
QY 262 CCTACAGTGTGTCAGCCACGCTTTATCCAGCATGAAGTGTGGAGAGATATAGCCACTG 321
DB 75 PROGLYARGVALYTHR-----LEUVALALAGLUTRPAIAPRO--- 86
QY 322 TTCATATGCCAAGGCTCGAGCCGACCTTGACGCCCTACCTGCTGATGGCTCACTTGTAT 381
DB 87 -----GIIIEYTHIASPHRSERARGPROALALEUENUEHISGLYHISERASP 103
QY 382 GTGGTCCCTGCCCCGTGAMAGGCTGGGAGGTGCCCATTCCTGTGGTGGAGCGTGAT 441
DB 104 THVALPROHEGLIALALASPRTRPHRHSHISPROLEUSERGLYGLIILEHISASP 123
QY 442 GCGGTATCTATNGTGGGGCACACTGACGACAGACAGACTGTGTGATGGCACTTACTGCA 501
DB 124 ASNCYVALTRPGLYARGLYALALIASPMETLYGLYPHEULAMETVALLEUSER 143
QY 502 GCGTGGAGCTCTGCTGATACAGAGATACATCCCGGAAGCTTTCTTCATTTCTCTG 561
DB 144 ALALIEARGIALARGGLINARGGLYGLVALPROSERARGPROILEARGPHEILEKET 163
QY 562 GCGCATGATGAGATCATCAGGAGACA-----GGGCTCAGAGATC 603
DB 164 PHEALASPROGLINUCYSERGLYTHRLIENGLYSERTHRTIPLEUGLYALATHRHISPRO 183
QY 604 TCAGCCCTGCTACAGTCAAGGGGCTGCCAGCTAGCCTTATGTGGACGAGGGGGCTTC 663
DB 184 GLIALAPHE-----ASPLIYVALTHRGUALA---ILESERGLUVALGLIYGLYPHE 199
QY 664 ATCTGGATGATTTCACTTCTTACTCATCAAGAGCCCATGCGCTTGATTCAGTCTCAG 723
DB 200 SERLEUTHR-----THRPROGLINGLYLSARG--VALTYVALIIEGLINSERIALAGLN 216
QY 724 AAGGTTCCATGACCTCAATGCTGCAAGTAAACATGATGATTCAGGCCACTTCAGCTCCT 783
DB 217 LYSGLYLEUTHRTIPRPHARGMETSERIALATHRGLYSERTHRCILYHISGLYSERMETARG 236
QY 784 CCAAGAGACAAAGCAFTGGCATCTTGCAAGCTGCTGCAAGCCGATTCGACAGACACACA 843
DB 237 ASPNPROASPASNALAVALTHARGVALLEUASPLALEUSERARGILEASPERTYGLIN 256
QY 844 ATGCGTATCATATTGGAGCGGAGACAGTGTACTGTATTCGACGCACTGGCAATGAG 903
DB 257 TRPPIROASPLEUHHIS-----PROVALIENGLINLUPHELEUENAGLIN 271
QY 904 TTTCCCTTCCCTGTCAATATATCTGACCAACCCATGCTATTTGAACCACTTATATAGC 963
DB 272 VALALALAMETTRPGLYLEUTHRIIEASPARGLASPLEU---GLUSERSERLEUSER 290
QY 964 AGGTTTATGAGAGAAATCCCTTAACCAATGCATATATCAGACACACAGGCACTCACC 1023
DB 291 PROILEGLYSERLEUSERARGMETVALALALACYSALAHISASNVALTHRPROTHR 310
QY 1024 ATATTCAAGCAGGGGCAATGCATATGTCATCCCGCCAGTGGCCGACCAAGTCAAC 1083
DB 311 VALLEUSERALAGLYTRYLYSVALASNVALVALPROTHRARGALASERIALAGLUALASP 330
  
```

```

QY 1084 TTCGGATTCACCTTGACAGACAGTCCAGAGGTCTAGAACTCAGGAAACATTGTG 1143
DB 331 ALAARGPHEILEPROGLY-----ALAGLINLUSPMETILLEUTHRIELYSERLEU 348
QY 1144 GCTGATACAGATTCACATTCATCTGTGTGAGTGCCTTTGACCCCTCCCGTACGCCCT 1203
DB 349 ALAGLYPROGLYILEASPPHEGLUTHRILESERARG--LYSPROALATHRALIALAPRO 367
QY 1204 TCTGATGACAAAGCCCTTGAGCTACACGCTGCTCCGACAGCCGATACAGATCCGCTTCCG 1263
DB 368 PHEGLINGLYALALALAVEL-----ASPLALIEARGATGALAVLALASPLALIASP 385
QY 1264 GAAGTCATATTTACTGCCCCAGTTACTTCTATTTGGCAACAGACAGACGCAATTCATA 1333
DB 386 GLYALA---VALVALLEUPROTYRILEUSNSERIALAGLYTHRASPALALYSGLYPHEALA 404
QY 1324 AACCTCACCACTGGC-----ATCTACAGGTTCTACCCCATCTACATACAGCT 1371
DB 405 VALLEUPROASPLIYARGARGILEASNTRYRGLYCSTHTRPROLEUARGLEU---PRO 423
QY 1372 GAAGACTTCAAACGCATC-----CATGGAGTCAACGAGAAATCTCAGTCCAAAGCC 1432
DB 424 ALASPPHEASPPHEILEASNLEUPHEHISGLYVALASPLIUALARGVALPROVALGLYSER 443
QY 1423 TATGAGACCAAGTGAAATTCATCTTTGAGTGTGATTCAGAAATGCT 1467
DB 444 LEUVALIPHEGLYALALYSVALVALASPHISILEUENGLINLUALA 458
  
```

Search completed: June 27, 2003, 11:23:51
 Job time : 175.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleole - nucleole search, using sw model

Run on: July 1, 2003, 05:52:17 ; Search time 3668 Seconds
(without alignments)
10343.515 Million cell updates/sec

Title: US-10-014-896-1

Sequence: 1 atgagctcagcggcgtcttg.....ctcactgcacaaactgtga 1509

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```

Pending_Patents_NA_Main:
1: /cgn2_6/ptodata/1/pna/US0900_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US0901_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US0902_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US0903_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US0904_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US0905_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US0906_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US0907_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US0908_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US0909_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US0910_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US0911_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US0912_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US0913_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US0914_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US0915_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US0916_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US0917_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US0918_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US0919_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US0920_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US0921_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US0922_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US0923_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US0924_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US0925_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US0926_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US0927_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US0928_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US0929_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US0930_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US0931_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US0932_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US0933_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US0934_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US0935_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US0936_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US0937_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US0938_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US0939_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US0940_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US0941_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US0942_COMB.seq.*

```

Result No.	Score	% Match	Query Length	DB ID	Description
1	1509	100.0	1509	US-10-014-896-1	Sequence 1, App11
2	1505.8	99.8	1509	US-60-345-785-3	Sequence 3, App11
3	1505.8	99.8	1509	US-60-345-785-1	Sequence 1, App11
4	1014	67.2	1086	US-10-014-896-3	Sequence 3, App11
5	860.4	57.0	932	US-09-648-365-17	Sequence 17, App1
6	860.4	57.0	932	US-10-259-776-17	Sequence 17, App1
7	673.4	44.6	1271	US-09-656-026-75	Sequence 75, App1
8	673.4	44.6	1271	US-60-156-623-265	Sequence 265, App1
9	596.8	39.5	909	US-60-360-207-413	Sequence 413, App1
10	582	38.6	1307	PCT-US02-09921-430	Sequence 430, App1
11	572.8	38.0	722	US-09-648-365-18	Sequence 18, App1
12	572.8	38.0	722	US-10-259-776-18	Sequence 18, App1
13	473.4	31.4	649	US-60-213-362-6769	Sequence 6769, App1
14	473.4	31.4	649	US-60-250-830-1230	Sequence 1230, App1
15	473.4	31.4	649	US-60-323-966-1230	Sequence 1230, App1
16	403.6	26.7	706	US-60-213-843-615	Sequence 615, App1
17	359.4	23.8	2432	US-60-360-207-412	Sequence 412, App1
18	312	20.7	312	US-09-540-733-6210	Sequence 6210, App1
19	297	19.7	308	US-09-540-733-5997	Sequence 5997, App1
20	285.2	18.9	391	US-60-252-833-40714	Sequence 40714, App1
21	267	17.7	319	US-08-904-468-45	Sequence 45, App1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Db	481	TCTGTGATGGCATTACTGGAGGCGCTTGGAGCTCTGTCTGATGAGAAATGATACCCCGCA	540
QY	541	AGATCTTCTTCATTTCTCTGGGCCATGATGAGAGTCAATCAGGGACAGGGGCTCAGAG	600
Db	541	AGATCTTCTTCATTTCTCTGGGCCATGATGAGAGTCAATCAGGGACAGGGGCTCAGAG	600
QY	601	ATCTCAGCCCTGCTACAGTCAGAGGGGGGTCCAGCTAGCCCTCAATTTGTGACGAGGGGGG	660
Db	601	ATCTCAGCCCTGCTACAGTCAGAGGGGGGTCCAGCTAGCCCTCAATTTGTGACGAGGGGGG	660
QY	661	TTTCATCTTGGATGATTTTCATTTCTCTAACTTCAGAAAGCCATGCGCTTGATTCAGTCTCA	720
Db	661	TTTCATCTTGGATGATTTTCATTTCTCTAACTTCAGAAAGCCATGCGCTTGATTCAGTCTCA	720
QY	721	GAGAGGGGTCCATGAAACCTCATGCTCAGTGAATAACTGACTTCAGGCCACTCTTCAGCT	780
Db	721	GAGAGGGGTCCATGAAACCTCATGCTCAGTGAATAACTGACTTCAGGCCACTCTTCAGCT	780
QY	781	CCCTCAAGAGGAGACAGCATTTGGCATCTTGACGCTCTGACGCTGTCAGCCGATTTGGACAGACA	840
Db	781	CCCTCAAGAGGAGACAGCATTTGGCATCTTGACGCTCTGACGCTGTCAGCCGATTTGGACAGACA	840
QY	841	CCAATGCCATCATATTTTGGAGCGGGGACAGTGGTGACTGTATTTGACGAACTGGGCAAT	900
Db	841	CCAATGCCATCATATTTTGGAGCGGGGACAGTGGTGACTGTATTTGACGAACTGGGCAAT	900
QY	901	GAGTTTCCCTTCCCTGTCATATATATCTGAGCAACCCATGGCTATTTGAACCACTATA	960
Db	901	GAGTTTCCCTTCCCTGTCATATATATCTGAGCAACCCATGGCTATTTGAACCACTATA	960
QY	961	AGCAGGTTTATGAGAGAAATCCCTTAACCAATGCATATATCAGAGCCACACGGGCACTC	1020
Db	961	AGCAGGTTTATGAGAGAAATCCCTTAACCAATGCATATATCAGAGCCACACGGGCACTC	1020
QY	1021	ACCATATTCAAAGCAGGGGTCAAGTCAATGTCATCCCGCCAGGGGCCAGGGCAGAGTC	1080
Db	1021	ACCATATTCAAAGCAGGGGTCAAGTCAATGTCATCCCGCCAGGGGCCAGGGCAGAGTC	1080
QY	1081	AACCTCCGAGATTCACCCCTGGAGACAGACGTCCAAAGAGGTCTTGAAGCTCAGAGAACATT	1140
Db	1081	AACCTCCGAGATTCACCCCTGGAGACAGACGTCCAAAGAGGTCTTGAAGCTCAGAGAACATT	1140
QY	1141	GTGGGTGATTAACAGAGTCCAGTTCATGTGATGAGTGGCTTTGAGACCCCTCCCGTCAGC	1200
Db	1141	GTGGGTGATTAACAGAGTCCAGTTCATGTGATGAGTGGCTTTGAGACCCCTCCCGTCAGC	1200
QY	1201	CCCTTGATGACAAAGCCCTTGGGCTACAGAGTGTCCGACAGACCGTACAGTCCGTCTTC	1260
Db	1201	CCCTTGATGACAAAGCCCTTGGGCTACAGAGTGTCCGACAGACCGTACAGTCCGTCTTC	1260
QY	1261	CCGGAAGTCAATATTAATGCCCCCAATTTACTTATTTGGCAACAGACAGCGGATTTCTTT	1320
Db	1261	CCGGAAGTCAATATTAATGCCCCCAATTTACTTATTTGGCAACAGACAGCGGATTTCTTT	1320
QY	1321	ACAAACCTCACCACTGGCATCTACAGAGTTCATACCCCATCTACATACAGGCTGAAGACTTC	1380
Db	1321	ACAAACCTCACCACTGGCATCTACAGAGTTCATACCCCATCTACATACAGGCTGAAGACTTC	1380
QY	1381	AAACGCATCATGAGAGTCAAGAGAAAATCTAGTCCAAAGCCTTATGAGCCCAAGTGA	1440
Db	1381	AAACGCATCATGAGAGTCAAGAGAAAATCTAGTCCAAAGCCTTATGAGCCCAAGTGA	1440
QY	1441	TTTCATCTTGGATGATGATTCAGAAATGCTGACACAGACAGAGGACCAAGTTTCTACCTGCAC	1500
Db	1441	TTTCATCTTGGATGATGATTCAGAAATGCTGACACAGACAGAGGACCAAGTTTCTACCTGCAC	1500
QY	1501	AAACTGTGA 1509	
Db	1501	AAACTGTGA 1509	

APPLICANT: LO, Audrey
APPLICANT: LAN, Ruth Y.
APPLICANT: URASHKA, Michael
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1232 PCT
CURRENT APPLICATION NUMBER: PCT/US02/09921
PRIORITY FILING DATE: 2002-03-27
PRIORITY FILING DATE: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
60/291,829; 60/299,428; 60/300,001; 60/299,776
PRIORITY FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
NUMBER OF SEQ ID NOS: 1146
SOFTWARE: PERL Program
SEQ ID NO: 430
LENGTH: 1307
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incycle ID No: LG:369881.5:2001JUN22
PCT-US02-09921-430

Query Match 38.6%; Score 582; DB 1; Length 1307;

Best Local Similarity 89.0%; Pred. No. 2,4e-159; Indels 85; Gaps 2;

Matches 687; Conservative 0; Mismatches 0;

QY 761 CTTCAGGCGCCTCTTCAGCTCTCCCAAGAGACAGCATTGGCATCTTCAGCTGCTG 820
DB 2 CTTCAGGCGCCTCTTCAGCTCTCCCAAGAGACAGCATTGGCATCTTCAGCTGCTG 61
QY 821 TCAGCCGATTTGGAGACAGACACCATTCCTATCATTTTGGAAAGGGGACATGGTGA 880
DB 62 TCAGCCGATTTGGAGACAGACACCATTCCTATCATTTTGGAAAGGGGACATGGTGA 121
QY 881 TATTGACAGCACTGGCAATGATTTCCTTCCTTCCTATCATTTTGGAAAGGGGACAT 940
DB 122 TATTGACAGCACTGGCAATGATTTCCTTCCTTCCTATCATTTTGGAAAGGGGACAT 142
QY 941 GGCATTATTTAAACCACTTTAAAGCAGTTTATGGAGAGAAATCCCTTAACCAATGCA 1000
DB 143 GGCATTATTTAAACCACTTTAAAGCAGTTTATGGAGAGAAATCCCTTAACCAATGCA 179
QY 1001 TCAGGACCGACCGGACCTACACATTTCAAGAGCGGGGTCA 1042
DB 180 TCAGGACCGACCGGACCTACACATTTCAAGAGCGGGGTCAAGGGGTTTGTGCTTC 239
QY 1043 -----AGTTCAATGTCATCCCCCGAGTGGCCAGGCCACAGTCAATTCGGGATTCA 1097
DB 240 TCTCTAGTTCAATGTCATCCCCCGAGTGGCCAGGCCACAGTCAATTCGGGATTCA 299
QY 1098 TGGACAGACAGTCGAAGAGTCTTGAATCTACGAAGAAATGTCGTCATTAACAGAGT 1157
DB 300 TGGACAGACAGTCGAAGAGTCTTGAATCTACGAAGAAATGTCGTCATTAACAGAGT 359
QY 1158 CCAGTTTCATGTTGAGTGCCTTTGACCCCTCCCGTCAGCCCTTCGATGACAAAGC 1217
DB 360 CCAGTTTCATGTTGAGTGCCTTTGACCCCTCCCGTCAGCCCTTCGATGACAAAGC 419
QY 1218 CTTGGGCTACAGCTGCTCCGCAAGCCTACAGTCCGTTCCCGGAAAGTAAATTTAC 1277
DB 420 CTTGGGCTACAGCTGCTCCGCAAGCCTACAGTCCGTTCCCGGAAAGTAAATTTAC 479
QY 1278 TGGCCGATTTCTATTTGGAACACAGACAGCGGATTTTACAAACCTACACAGT 1337
DB 480 TGGCCGATTTCTATTTGGAACACAGACAGCGGATTTTACAAACCTACACAGT 539
QY 1338 CATCTACAGTTCTACCCCATCTACATACAGCTGAAAGCTCAAGCATCATGAGT 1397
DB 540 CATCTACAGTTCTACCCCATCTACATACAGCTGAAAGCTCAAGCATCATGAGT 599
QY 1398 CAACGAGAAATCTCAGTCCAGAGCTATAGAGCCCAAGTGAATTTATTTGAGTTGAT 1457
DB 600 CAACGAGAAATCTCAGTCCAGAGCTATAGAGCCCAAGTGAATTTATTTGAGTTGAT 659

QY 1458 TCAGATGCTGACACAGACAGCAGCAGCAGTTCCTACCTGCAACAACTGTGA 1509
DB 660 TCAGATGCTGACACAGACAGCAGCAGCAGTTCCTACCTGCAACAACTGTGA 711

RESULT 11

US-09-648-365-18

; Sequence 18, Application US/09648365

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding

; TITLE OF INVENTION: Human Protease-Like Homologs

; FILE REFERENCE: 5800-165

; CURRENT APPLICATION NUMBER: US/09/648,365

; PRIORITY FILING DATE: 2000-08-25

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO: 18

; LENGTH: 722

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-648-365-18

Query Match 38.0%; Score 572.8; DB 25; Length 722;

Best Local Similarity 98.8%; Pred. No. 9e-157; Indels 7; Gaps 0;

Matches 577; Conservative 0; Mismatches 7;

QY 257 TCTTCTACAGTGGTCAGCAGCAGCTTTATTCAGCATGAAAGTGTGGAAAGATATAGCC 316
DB 1 TCTTCTACAGTGGTCAGCAGCAGCTTTATTCAGCATGAAAGTGTGGAAAGATATAGCC 60
QY 317 ACCTGTCTATTCAGCAGCAGCTTCGACCCCGCTTGGAGCCCTACGTCGATGGCTCAT 376
DB 61 ACCTGTCTATTCAGCAGCAGCTTCGACCCCGCTTGGAGCCCTACGTCGATGGCTCAT 120
QY 377 TTGATGTCGTCCTGTCCTCCCTGAAGAGGCTGGAGAGTGGCCCATCTCTGAGTTGAGC 436
DB 121 TTGATGTCGTCCTGTCCTCCCTGAAGAGGCTGGAGAGTGGCCCATCTCTGAGTTGAGC 180
QY 437 GTGATGTCGTCCTGTCCTCCCTGAAGAGGCTGGAGAGTGGCCCATCTCTGAGTTGAGC 496
DB 181 GTGATGTCGTCCTGTCCTCCCTGAAGAGGCTGGAGAGTGGCCCATCTCTGAGTTGAGC 240
QY 497 TGCAGGCTGGAGCTCTCTGTCGATCAGAAATCATCCCGCAAGATCTTCTCATTT 556
DB 241 TGCAGGCTGGAGCTCTCTGTCGATCAGAAATCATCCCGCAAGATCTTCTCATTT 300
QY 557 CTCTGGGCGATGATGAGAGTCTACAGGAGCAGGGGCTCAGAGATCTCAGCCCTGCTAC 616
DB 301 CTCTGGGCGATGATGAGAGTCTACAGGAGCAGGGGCTCAGAGATCTCAGCCCTGCTAC 360
QY 617 AGTCAAGGGGCTCAGCTGACCTTCATTTGAGAGGAGGGGCTTCATCTTGGATGAT 676
DB 361 AGTCAAGGGGCTCAGCTGACCTTCATTTGAGAGGAGGGGCTTCATCTTGGATGAT 420
QY 677 TCATTCCTAATTCAGAAAGCCATCGCTTATTTGAGAGTCTCAGAAAGGTTTCATGA 736
DB 421 TCATTCCTAATTCAGAAAGCCATCGCTTATTTGAGAGTCTCAGAAAGGTTTCATGA 480
QY 737 ACCTCAGCTCAGAAAGCCATCGCTTATTTGAGAGTCTCAGAAAGGTTTCATGA 796
DB 481 ACCTCAGCTCAGAAAGCCATCGCTTATTTGAGAGTCTCAGAAAGGTTTCATGA 540
QY 797 GCATGGCATCTTGCAGCTGCTGTCAGCGATTTGAGAGCA 840
DB 541 GCATGGCATCTTGCAGCTGCTGTCAGCGATTTGAGAGCA 584

RESULT 12

US-10-259-776-18

; Sequence 18, Application US/10259776

; GENERAL INFORMATION:

```
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding
FILE REFERENCE: 5800-165
CURRENT APPLICATION NUMBER: US/10/259,776
CURRENT FILING DATE: 2002-09-27
PRIOR APPLICATION NUMBER: US/09/648,365
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 18
LENGTH: 722
TYPE: DNA
ORGANISM: Homo sapiens
US-10-259-776-18
```

```
Query Match      38.0%; Score 572.8; DB 42; Length 722;
Best Local Similarity 98.8%; Pred. No. 9e-157;
Matches 577; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
OY 257 TCTTCTACAGTGTGACACACAGCTTTATCCAGCATGAAGTCGTGGAAGATAGCC 316
DB 1 TCTTCTACAGTGTGACACACAGCTTTATCCAGCATGAAGTCGTGGAAGATAGCC 60
OY 317 ACCTGTCATCATCAAGAGCTCGAGCCAGCTTCAGCCCTACCTGCTGATGCTCAGT 376
DB 61 ACCTGTCATCATCAAGAGCTCGAGCCAGCTTCAGCCCTACCTGCTGATGCTCAGT 120
OY 377 TTGATGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436
DB 121 TTGATGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
OY 437 GTGATGCGTATGATGTCGAGGACACTGAGCAGCAAGAACTGTGATGCTATAC 496
DB 181 GTGATGCGTATGATGTCGAGGACACTGAGCAGCAAGAACTGTGATGCTATAC 240
OY 497 TGCAGGCTTGGAGCTCTGCTGATCAGAGATACATCCCCCAAGATCTTTTATTT 556
DB 241 TGCAGGCTTGGAGCTCTGCTGATCAGAGATACATCCCCCAAGATCTTTTATTT 300
OY 557 CTCGAGGCTGATGAGAGATCAGAGGAGGAGGCTCAGAGATCTAGCCCTGCTAC 616
DB 301 CTCGAGGCTGATGAGAGATCAGAGGAGGAGGCTCAGAGATCTAGCCCTGCTAC 360
OY 617 AGTCAAGGCGCTCCAGCTAGCTTCAATGTGACGAGGAGGCTTCAATTTGATGAT 676
DB 361 AGTCAAGGCGCTCCAGCTAGCTTCAATGTGACGAGGAGGCTTCAATTTGATGAT 420
OY 677 TCATTCCTACATCAAGAGCCCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 736
DB 421 TCATTCCTACATCAAGAGCCCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 480
OY 737 ACCTCATCTCGAAGTAACATGACTTCAGGCACTCTTCACTCTCCAAAGAGACAA 796
DB 481 ACCTCATCTCGAAGTAACATGACTTCAGGCACTCTTCACTCTCCAAAGAGACAA 540
OY 797 GCATGGCATCTTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 541 GCATGGCATCTTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 584
```

```
RESULT 13
US-60-213-362-6769/C
; Sequence 6769, Application US/60213362
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
; FILE REFERENCE: GX-0016 P
; CURRENT APPLICATION NUMBER: US/60/213,362
```

```
CURRENT FILING DATE: 2000-06-21
NUMBER OF SEQ ID NOS: 8429
SOFTWARE: PERL Program
SEQ ID NO 6769
LENGTH: 649
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 423998.2
US-60-213-362-6769
```

```
Query Match      31.4%; Score 473.4; DB 65; Length 649;
Best Local Similarity 96.6%; Pred. No. 1.4e-127;
Matches 488; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
```

```
OY 1 ATGGCTCAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
DB 562 ATGGCTCAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
OY 61 ACCGCTTCAGATGATGAGGCGCGAGAGAGCGGAGCATCAAGGCGTCCGAATCCT 120
DB 502 ACCGCTTCAGATGATGAGGCGCGAGAGAGCGGAGCATCAAGGCGTCCGAATCCT 443
OY 121 TCTCAGTTCAGCAAGAGAGAAAGCGTCCGATGAAGAGAGCGCTGAAGAGTCCATCCAG 180
DB 442 TCTCAGTTCAGCAAGAGAGAAAGCGTCCGATGAAGAGAGCGCTGAAGAGTCCATCCAG 383
OY 181 ATTCAACAGTATGATTTTACCTGAGAAATCCAAATCTACAGCTGCTGATGCTGGA 240
DB 382 ATTCAACAGTATGATTTTACCTGAGAAATCCAAATCTACAGCTGCTGATGCTGGA 323
OY 241 AAATACATTCATAAGCTTTCTCTACAGTGTGAGCAGCAGCTTTATCCAGATGAATC 300
DB 322 AAATACATTCATAAGCTTTCTCTACAGTGTGAGCAGCAGCTTTATCCAGATGAATC 263
OY 301 GTGGAAGATATACCACTGTTTACATCAATCAAGGCTCGAGCCCAAGCTTTCAGCCCTAC 360
DB 262 GTGGAAGATATACCACTGTTTACATCAATCAAGGCTCGAGCCCAAGCTTTCAGCCCTAC 204
OY 361 CTGCTGATGCTCACTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 203 CTGCTGATGCTCACTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144
OY 421 TTCTGTGGTGTGAGCGTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 480
DB 143 TTCTGTGGTGTGAGCGTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 84
OY 481 TCTGTGATGCTATTA 495
DB 83 TCTGTGATGCTCTGA 69
```

```
RESULT 14
US-60-250-830-1230/C
; Sequence 1230, Application US/60250830
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
; FILE REFERENCE: GX-0020 P
; CURRENT APPLICATION NUMBER: US/60/250,830
; CURRENT FILING DATE: 2000-11-04
; NUMBER OF SEQ ID NOS: 3246
; SOFTWARE: PERL Program
; SEQ ID NO 1230
; LENGTH: 649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 42398.2
US-60-250-830-1230

Query Match 31.4%; Score 473.4; DB 69; Length 649;
Best Local Similarity 98.6%; Pred. No. 1.4e-127;
Matches 488; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
QY 1 ATGGCTCAGCGGTGGGCTTTGGCTGGCCCTGGTGGCTATGCTGCTCTAGTTTCCCT 60
DB ATGGCTCAGCGGTGGGCTTTGGCTGGCCCTGGTGGCTATGCTGCTCTAGTTTCCCT 503
QY 61 ACCGTCTCAGATGATGAGGCGCCGAGAGAGCGGGAGCATCAAGGGCGTGGGAATCCCT 120
DB ACCGTCTCAGATGATGAGGCGCCGAGAGAGCGGGAGCATCAAGGGCGTGGGAATCCCT 443
QY 121 TCTCAGTTACGCAAGAGAGAGAGCGCGTGGGATGAAAGAGCGCTGAAAGTGCATCCAG 180
DB 442 TCTCAGTTACGCAAGAGAGAGAGCGCGTGGGATGAAAGAGCGCTGAAAGTGCATCCAG 383
QY 181 ATTCCAAGAGTACTTTAGCTCTGAGAAAGTCCAACTACAGCCCTGGCTGAGTTCGGA 240
DB 382 ATTCCAAGAGTACTTTAGCTCTGAGAAAGTCCAACTACAGCCCTGGCTGAGTTCGGA 323
QY 241 AATATCATTCATTAAGTCTTCTTCTACAGTGGTCCAGCAGCTTTATCCAGCATGAATC 300
DB 322 AATATCATTCATTAAGTCTTCTTCTACAGTGGTCCAGCAGCTTTATCCAGCATGAATC 263
QY 301 GTGGAAGATATAGCCACCTGTTCACTATCCAAAGGCTCGAACCAGCTTGACAGCCCTAC 360
DB 262 GTGGAAGATATAGCCACCTGTTCACTATCCAAAGGCTCGAACCAGCTTGACAGCCCTAC 204
QY 361 CTGCTGATGCTCACTTTGATGTGTGCTGCTGCCCTGGAAGAGGCTGGAGTGCCTCA 420
DB 203 CTGCTGATGCTCACTTTGATGTGTGCTGCTGCCCTGGAAGAGGCTGGAGTGCCTCA 144
QY 421 TTCTCTGGGTGGAGCGGTGATGCGCTCATCTATGCTGGGGCAGCACTGACGCAAGAAC 480
DB 143 TTCTCTGGGTGGAGCGGTGATGCGCTCATCTATGCTGGGGCAGCACTGACGCAAGAAC 84
QY 481 TCTGTGATGGCATTA 495
DB 83 TCTGTGATGGCTTGA 69
```

RESULT 15
US-60-323-966-1230/c
Sequence 1230, Application US/60323966
GENERAL INFORMATION:
APPLICANT: Morris, MacDonald
APPLICANT: Lal, Preeti
APPLICANT: Diep, Dinh
TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
TITLE OF INVENTION: POLYMORPHISM SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
FILE REFERENCE: GX-0020-1 P
CURRENT APPLICATION NUMBER: US/60/323,966
CURRENT FILING DATE: 2001-09-21
NUMBER OF SEQ ID NOS: 3246
SOFTWARE: PERL Program
SEQ ID NO 1230
LENGTH: 649
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 42398.2
US-60-323-966-1230

Query Match 31.4%; Score 473.4; DB 76; Length 649;
Best Local Similarity 98.6%; Pred. No. 1.4e-127;
Matches 488; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
QY 1 ATGGCTCAGCGGTGGGCTTTGGCTGGCCCTGGTGGCTATGCTGCTCTAGTTTCCCT 60
DB ATGGCTCAGCGGTGGGCTTTGGCTGGCCCTGGTGGCTATGCTGCTCTAGTTTCCCT 503
QY 61 ACCGTCTCAGATGATGAGGCGCCGAGAGAGCGGGAGCATCAAGGGCGTGGGAATCCCT 120
DB ACCGTCTCAGATGATGAGGCGCCGAGAGAGCGGGAGCATCAAGGGCGTGGGAATCCCT 443
QY 121 TCTCAGTTACGCAAGAGAGAGAGCGCGTGGGATGAAAGAGCGCTGAAAGTGCATCCAG 180
DB 442 TCTCAGTTACGCAAGAGAGAGAGCGCGTGGGATGAAAGAGCGCTGAAAGTGCATCCAG 383
QY 181 ATTCCAAGAGTACTTTAGCTCTGAGAAAGTCCAACTACAGCCCTGGCTGAGTTCGGA 240
DB 382 ATTCCAAGAGTACTTTAGCTCTGAGAAAGTCCAACTACAGCCCTGGCTGAGTTCGGA 323
QY 241 AATATCATTCATTAAGTCTTCTTCTACAGTGGTCCAGCAGCTTTATCCAGCATGAATC 300
DB 322 AATATCATTCATTAAGTCTTCTTCTACAGTGGTCCAGCAGCTTTATCCAGCATGAATC 263
QY 301 GTGGAAGATATAGCCACCTGTTCACTATCCAAAGGCTCGAACCAGCTTGACAGCCCTAC 360
DB 262 GTGGAAGATATAGCCACCTGTTCACTATCCAAAGGCTCGAACCAGCTTGACAGCCCTAC 204
QY 361 CTGCTGATGCTCACTTTGATGTGTGCTGCTGCCCTGGAAGAGGCTGGAGTGCCTCA 420
DB 203 CTGCTGATGCTCACTTTGATGTGTGCTGCTGCCCTGGAAGAGGCTGGAGTGCCTCA 144
QY 421 TTCTCTGGGTGGAGCGGTGATGCGCTCATCTATGCTGGGGCAGCACTGACGCAAGAAC 480
DB 143 TTCTCTGGGTGGAGCGGTGATGCGCTCATCTATGCTGGGGCAGCACTGACGCAAGAAC 84
QY 481 TCTGTGATGGCATTA 495
DB 83 TCTGTGATGGCTTGA 69
```

Search completed: July 1, 2003, 08:49:56
Job time : 3689 secs


```
Db 723 TTCATCTTGATGATGATTTTCAATCTTAACTCAAGAACCCATGCCCTTGATGCACTCA 782
QY 721 GAGAGGGTTCCATGAACTCATGCTGCAAGTAACATATCTTCAAGCCACCTTCAGCT 780
Db 783 GAGAGGGTTCCATGAACTCATGCTGCAAGTAACATATCTTCAAGCCACCTTCAGCT 842
QY 781 CCTCCAAAGAGACAGATTTGGCATCTTGGCAGCTGCTGCAAGCCATTTGAGAGACA 840
Db 843 CTTCCAAAGAGACAGATTTGGCATCTTGGCAGCTGCTGCAAGCCATTTGAGAGACA 902
QY 841 CCAATGCTATCATTTTGGAAACGGGACAGTGCTGATGTATGACAGCAATGCGAAT 900
Db 903 CCAATGCTATCATTTTGGAAACGGGACAGTGCTGATGTATGACAGCAATGCGAAT 962
QY 901 GATTTCCCTTCCCTTCAATATATCTGAGAACCCATGGCTATTTGAACTTATA 960
Db 963 G----- 963
QY 961 AGCAGTTTATGAGAGAAATCCCTTAACCAATGCAATATAGAGACACAGGACACTC 1020
Db 964 ---AGTTTATGAGAGAAATCCCTTAACCAATGCAATATAGAGACACAGGACACTC 1020
QY 1021 ACCATATTCAGAGAGGGGTCAAGTTCATGTATGCTATGCTGCGCCAGGCGCACAGTC 1080
Db 1021 ACCATATTCAGAGAGGGGTCAAGTTCATGTATGCTATGCTGCGCCAGGCGCACAGTC 1080
QY 1081 AACTTCGGATTCACCTTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140
Db 1081 AACTTCGGATTCACCTTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140
QY 1141 GTGGCTGATTAACAGAGTCCAGTTCATGTATGAGTCCCTTGAACCCCTCCCGTCAGC 1200
Db 1141 GTGGCTGATTAACAGAGTCCAGTTCATGTATGAGTCCCTTGAACCCCTCCCGTCAGC 1200
QY 1201 CCTTCTGATGACAGAGGCTTGGGTACAGAGTGTGCTGCGCCAGAGAGAGAGAGAGAT 1260
Db 1201 CCTTCTGATGACAGAGGCTTGGGTACAGAGTGTGCTGCGCCAGAGAGAGAGAGAT 1260
QY 1261 CCGGAAGTCAATATTAATGCCCCAGTACTTCTTATTTGGCAACAGAGAGAGAGAT 1320
Db 1261 CCGGAAGTCAATATTAATGCCCCAGTACTTCTTATTTGGCAACAGAGAGAGAGAT 1320
QY 1321 ACAAACTGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Db 1321 ACAAACTGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1380
QY 1381 AAAGCATTCATGGA 1395
Db 1381 AAAGCATTCATGGA 1395
```

RESULT 7

PCT-US02-32850A-62

```
Sequence 62, Application PC/TUS0232850A
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: GORVAD, Ann E.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: EMERLING, Brooke M.
APPLICANT: YANG, Junning
APPLICANT: LEE, Soo Yeun
APPLICANT: TRAN, Uyen K.
APPLICANT: BECHA, Shanya D.
APPLICANT: DUGGAN, Brendan M.
APPLICANT: LEE, Ernestine A.
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: LI, Joana X.
APPLICANT: SPRAGUE, William W.
APPLICANT: HAFALIA, April J.A.
APPLICANT: CHAMLA, Narinder K.
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: KABLE, Amy E.
```

```
APPLICANT: YUE, Henry
APPLICANT: MARQUIS Joseph P.
APPLICANT: YAO, Monique G.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: TANG, Y. Tom
APPLICANT: JIN, Pei
APPLICANT: CHIEN, David
APPLICANT: BHATIA, Umesh G.
APPLICANT: BURRILL, John D.
APPLICANT: LEE, Sally
APPLICANT: BLAKE, Julie J.
APPLICANT: HO, Anne
APPLICANT: ZHENG, Wenjin
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1237 PCT
CURRENT APPLICATION NUMBER: PCT/US02/32850A
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: US 60/329,689
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/335,703
PRIOR FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: US 60/348,887
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: US 60/334,145
PRIOR FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: US 60/337,451
PRIOR FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: US 60/340,584
PRIOR FILING DATE: 2001-12-14
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PERL Program
SEQ ID NO 62
LENGTH: 1395
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 3118930CB1
PCT-US02-32850A-62
Query Match 83.5%; Score 1259.4; DB 1; Length 1395;
Best Local Similarity 95.5%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 1; Indels 62; Gaps 1;
QY 1 ATGGCTGAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Db 63 ATGGCTGAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
QY 61 ACCGTCACAGATGATGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 120
Db 123 ACCGTCACAGATGATGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 182
QY 121 TCTCAGTTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 180
Db 183 TCTCAGTTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 242
QY 181 ATTCCACAGTGAATTTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 240
Db 243 ATTCCACAGTGAATTTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 302
QY 241 AATATCATTAAGTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 300
Db 303 AATATCATTAAGTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 362
QY 301 GTGGAAGAGTATAGCAGCCTGTTCTATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAT 360
Db 363 GTGGAAGAGTATAGCAGCCTGTTCTATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAT 422
QY 361 CTGCTGATGCTCTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 423 CTGCTGATGCTCTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 482
QY 421 TTCTCTGGGTGAGAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
```

|||||
483 TTCTCTGGGTGGAGCGTGTATGGCTCATTAAGTGGGAGCAGCTGAGCAGACAGAAC 542
481 TCTGTATGATTAAGTCTGAGGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 540
543 TCTGTATGATTAAGTCTGAGGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 602
541 AGATCTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 600
603 AGATCTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 662
601 ATTCAGACCTCTGATCAGAGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 660
663 ATTCAGACCTCTGATCAGAGCTTGGAGCTCTCTGATCAGAGAAATATCCCGCA 722
661 TTCTATCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 720
723 TTCTATCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 782
721 GAGAGAGGTTCCATGAGACCTGATGAGTAAATGATGATGATGATGATGATGATGAT 780
783 GAGAGAGGTTCCATGAGACCTGATGAGTAAATGATGATGATGATGATGATGATGAT 842
781 CTTCAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
843 CTTCAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
841 CCATGCTCTCATATTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
903 CCATGCTCTCATATTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 962
901 GAGTTCCCTTCCCTGTCAATTAATTCATGAGCAGCAGCAGCAGCAGCAGCAGCAG 960
963 G----- 963
961 AGCAGTTTATGAGAGAAATCCCTTAACCATGATGATGATGATGATGATGATGATGAT 1020
964 ---AGTTTATGAGAGAAATCCCTTAACCATGATGATGATGATGATGATGATGATGAT 1020
1021 ACCATATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
1021 ACCATATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
1081 AACCTCGGATTCACCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1081 AACCTCGGATTCACCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1141 GTGGCTGATTAACAGAGTCCAGTTCATGATGATGATGATGATGATGATGATGATGAT 1200
1141 GTGGCTGATTAACAGAGTCCAGTTCATGATGATGATGATGATGATGATGATGATGAT 1200
1201 CCTTCTATGAGAGAGGCTTGGGCTTACAGAGCTCTCTGATGATGATGATGATGATGAT 1260
1201 CCTTCTATGAGAGAGGCTTGGGCTTACAGAGCTCTCTGATGATGATGATGATGATGAT 1260
1261 CCGAGAGTCAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
1261 CCGAGAGTCAATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
1321 ACAAACTGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
1321 ACAAACTGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
1381 AAAGCATCATGGA 1395
1381 AAAGCATCATGGA 1395
1381 AAAGCATCATGGA 1395

RESULT 8
US-10-144-771-413
; Sequence 413, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig

|||||
1 ATGGCTCAGGCGTGGCTTGGCTGCTGAGGCTTGGCTGATGATGATGATGATGATGATGAT 60
16 ATGGCTCAGGCGTGGCTTGGCTGCTGAGGCTTGGCTGATGATGATGATGATGATGATGAT 75
61 ACAGCTCAGGCGTGGCTTGGCTGCTGAGGCTTGGCTGATGATGATGATGATGATGATGAT 120
76 ACAGCTCAGGCGTGGCTTGGCTGCTGAGGCTTGGCTGATGATGATGATGATGATGATGAT 135
121 TTCTATCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 180
136 TTCTATCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 195
181 ATTCCAGAGTCTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 240
196 ATTCCAGAGTCTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 255
241 AAATCAATTCATTAATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 300
256 AAATCAATTCATTAATGATTTCTTCAATTTCTGAGGCTTGGAGCTCTCTGATCAGAG 303
301 GTGGAAGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
304 GTGGAAGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 363
361 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
364 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423
421 TTCTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGG 480
424 TTCTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGG 483
481 TCTGTATGATTAAGTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAG 540
484 TCTGTATGATTAAGTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAG 543
541 AGATCTTCTTCAATTTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAG 597
544 AGATCTTCTTCAATTTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAG 603
598 AGATCTCAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGG 657
604 AGATCTCAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGG 663
658 GCTTCAATCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGG 717
664 AGCTTCTTCTTGGATGATTTCTTCAATTTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGG 723
718 TCAAGAGAGGTTCCATGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
724 ACTGAGAGAGGTTCCATGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 783
778 GCTTCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 837
784 GCTTCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 843
838 ACACCATGCTTCAATTTCTGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAGGCTTGGAG 897

Db 844 ACACCAATGCCAATATGTTTGAGAGGGCCATTTGAAGACAAATCACTACTAGCA 903
QY 898 AATGAG 903
Db 904 AATGAG 909

RESULT 9

US-10-218-140-217
; Sequence 217, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/219,140
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curanator Version 1.0
; SEQ ID NO 217
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(499)
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: "n" = "a", "c", "t" or "g"
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (2)..(2)
; OTHER INFORMATION: "n" = "a", "c", "t" or "g"
US-10-218-140-217

Query Match 32.9%; Score 496.4; DB 9; Length 500;
Best Local Similarity 99.8%; Pred. No. 2,1e-134;
Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 141 ACGGTCGCGATGAAGAGCGCTGAAGGTGCCATCCAGATTCCAACTGACTTTAG 200
Db 3 ACGGTCGCGATGAAGAGCGCTGAAGGTGCCATCCAGATTCCAACTGACTTTAG 62
QY 201 CTCGAGAGATCCCAATCTACAGAGCCCTGGCTGAGTTGGAAATATCATTAAGCTT 260
Db 63 CTCGAGAGATCCCAATCTACAGAGCCCTGGCTGAGTTGGAAATATCATTAAGCTT 122
QY 261 TCCTACAGTGTGAGCAACAGCTTTATTCAGCATGAAGTGTGGAAGATATACCACT 320
Db 123 TCCTACAGTGTGAGCAACAGCTTTATTCAGCATGAAGTGTGGAAGATATACCACT 182
QY 321 GTTCACTATCCAAAGGCTGAGACCCAGCTTTCAGAGCCCTTACCTGCTGATGCTCACTTTGA 380
Db 183 GTTCACTATCCAAAGGCTGAGACCCAGCTTTCAGAGCCCTTACCTGCTGATGCTCACTTTGA 242
QY 381 TGTGTGCTGCTCCCTGAGAAAGGCTGAGAGTGTGAGAGTGTGAGAGGCTGGA 440
Db 243 TGTGTGCTGCTCCCTGAGAAAGGCTGAGAGTGTGAGAGTGTGAGAGGCTGGA 302
QY 441 TGGCTCATCTATGTTGGGCGACACTGAGACAGAAATCTGTGATGGCATTTACTGCA 500
Db 303 TGGCTCATCTATGTTGGGCGACACTGAGACAGAAATCTGTGATGGCATTTACTGCA 362

QY 501 GGCCCTTGAGAGCTCTGCTGATCAGAGATACATCCCCAGAGATCTTCTCATTTCTCT 560
Db 363 GGCCCTTGAGAGCTCTGCTGATCAGAGATACATCCCCAGAGATCTTCTCATTTCTCT 422
QY 561 GGCCCATATGAGAGTATCAGGAGACGGGGCTCAGAGATCTACACCCCTGCTACAGTC 620
Db 423 GGCCCATATGAGAGTATCAGGAGACGGGGCTCAGAGATCTACACCCCTGCTACAGTC 482
QY 621 AAGGGCGCTCCAGTACG 638
Db 483 AAGGGCGCTCCAGTACG 500

RESULT 10

US-10-144-771-412
; Sequence 412, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CLO01321
; CURRENT APPLICATION NUMBER: US/10/144,771
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 412
; LENGTH: 2432
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-412

Query Match 23.8%; Score 359.4; DB 9; Length 2432;
Best Local Similarity 77.0%; Pred. No. 5.8e-94;
Matches 438; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 941 GGCTATTGTAACCACTTATTAACAGGTTTATGAGAGAAATCCCTTAACATGCATTA 1000
Db 1 GGCTATTGTAACCACTTATTAACAGGTTTATGAGAGAAATCCCTTAACATGCATTA 60
QY 1001 TCAGAGACCAACAGGCTACCATATTTAAAGCAGGGGTCAAGTTCAATGATCCGCC 1060
Db 61 TCAGAGACCAACAGGCTACCATATTTAAATGCAAGGATCAAGGTGATGATCCCTC 120
QY 1061 CAGTGGCCAGGCGCAGTCAACTTCGGATTCACCTCGAGACAGAGTCAAGAGTCC 1120
Db 121 CATTGGCTCAGGCTCAATCAATCACTGCCGAATTCACCTTCGAGACAGTCAAGTCC 180
QY 1121 TAGAAGTCAAGAAACATTTGCTGATTAAGAGTCCAGTTCAGTGTGAGTGCCT 1180
Db 181 TAGAAGTCAAGAAACATTTGCTGATGAGAGAGTCCAGTTCAGTGTGAGTGCCT 240
QY 1181 TTGACCCCTCCCGCTCAGCCCTTCTGATGAGCAAGGCCCTTGGGCTACAGCTCCGCC 1240
Db 241 TTGACCCCTCCCGCTCAGCCCTTCTGATGAGCAAGGCCCTTGGGCTACAGCTCCGCC 300
QY 1241 AAGGCGTACAGTCCGCTTCCCGGAGTCAATTTACTCCCGAGTTACTTCTATTGGCA 1300
Db 301 AAGGCGTACAGTCCGCTTCCCGGAGTCAATTTACTCCCGAGTTACTTCTATTGGCA 1360
QY 1301 ACACAGACAGCCGATTTCTTACAAACCTCAACCACTGGATCAAGGTTCTACCCGATCT 1360
Db 361 ATAGGAGACAGCCGATTTCTTACAAACCTCAACCACTGGATCAAGGTTCTACCCGATCT 420
QY 1361 ACATPACAGCTGAGAGTCTTAAACGATCATGAGATCAACGAGAAATCTAGTCCAG 1420
Db 421 CCCTGAACCTCGAGAGTCTTAAAGTGTGATGAGTCAATGAGAAAGTTCCGTTCA 480
QY 1421 CATTATGAGCCCAAGTAAATTCATCTTTGAGTTGATGAGTGTGAGTGTGAGTGTGAG 1480
Db 481 ACTATGAGCCAGAGTAAATTCATCTTTGAGTTGATGAGTGTGAGTGTGAGTGTGAG 540
QY 1481 AGCCAGTTTCTCACTGACCAAACTGTGA 1509
Db 541 AGCCAGTTTCTCACTGACCAAACTGTGA 569

OY	181	ATTCACAAAGTACTTTTAGCTCTGAGAACTCAATTA	CTACAGCCCTGGATTCGGA	240
Db	185	-----	-----	184
OY	241	AAATACATTCATAAAGCTCTTTCCTACAGTGGTACACACAGCTTTATCCACAGTAAGTC		300
Db	185	-----	AAAGTCTTTCCTACAGTGGTACACACACACTTTATCCACAGTAAGTC	232
OY	301	GTGGAAGAGTATAGCCACCTGTTCTACTATCCAAAGGCTCGAGCCCACTTGACAGCCCTAC		360
Db	233	GTGGAAGAGTATAGCCACCTGTTCTACTATCCAAAGGCTCGAGCCCACTTGACAGCCCTAC		292
OY	361	CTGCTATAGGCTCACTTTGATGTGGTGCCTGACCCCTGAAGAAAGGCTGGAGGTGCCCA		420
Db	293	CTGCTATAGGCTCACTTTGATGTGGTGCCTGACCCCTGAAGAAAGGCTGGAGGTGCCCA		352
OY	421	TTCTCTGGG	429	
Db	353	TTCTCTGGG	361	

```

RESULT 13
US-09-947-907-1617/c
; Sequence 1617, Application US/09947907
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: COLLECTION OF CODING REGION SINGLE
; TITLE OF INVENTION: NUCLEOTIDE POLYMORPHISM (CSNPS) LOCATED ON EACH OF THE HUMAN
; TITLE OF INVENTION: CHROMOSOMES, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CLO00896
; CURRENT APPLICATION NUMBER: US/09/947, 907
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 21266
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1617
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-947-907-1617

```

Query Match	15.7%;	Score 236.4;	DB 7;	Length 601;
Best Local Similarity	96.8%;	Pred. No. 3.1e-58;		
Matches 240; Conservative	1;	Mismatches 7;	Indels 0;	Gaps 0

OY	248	TTATATAAGCTTTCTCTACAGTGGTCAGACACAGTTATCTACATGAATGCTGGAG	307
	498	TGCTTCAAGCTTTCTCTACAGTGGTCAGACACAGTTATCTACAGATGAATGCTGGAG	439
OY	308	AGTATAGCCACTGTTCACTATTCAGAGGCTGGACCCAGCTTGACGCTTACTGCTGA	367
	438	AGTATAGCCACTGTTCACTATTCAGAGGCTGGACCCAGCTTGAGGCTTACTGCTGA	379
Db	368	TGGCTACTTAAATGTGTGCTGCTGGCCCTGAAGAAGGCTGGGAGGTGCCCATTTCTGTG	427
	378	TGGCTACTTAAATGTGTGCTGCTGGCCCTGAAGAAGGCTGGGAGGTGCCCATTTCTGTG	319
OY	428	GATTGAGCGTGATGGCGTCAATCTAATGGTTCGGGGCACACTGGACAGCAAGACTCTGTGA	487
Db	318	GATTGAGCGTGATGGCGTCAATCTAATGGTTCGGGGCACACTGGACAGCAAGACTCTGTGA	259
OY	488	TGGCATTA	495
Db	258	TGGTGTGA	251

RESULT 14
US-60-452-680-42976

```

; Sequence 42976, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GROPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42976
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-42976

```

Query Match	13.3%	Score 200.6;	DB 12;	Length 201;
Best Local Similarity	99.5%;	Pred. No. 5.7e-48;		
Matches 200; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

OY	1034	CAGGGGCTAAAGTTCAATGCATCAATCCCCCAGTGGCCAGAGCCACACATCAACTTCGGGATTC	103
Db	1	CAGGGGCTAAAGTTCAATGCATCAATCCCCCAGTGGCCAGAGCCACACATCAACTTCGGGATTC	60
OY	1094	ACCCGTGACAGACAGTCCACAGAGTCTCTAGAACTACAGAAAGAACTTTGTGGCTGATTAACA	115
Db	61	ACCCGTGACAGACAGTCCACAGAGTCTCTAGAACTACAGAAAGAACTTTGTGGCTGATTAACA	120
OY	1154	GAGTTCAGTTCATATGTGTGAGTGGCTTTTGACCCCTTCCCGCTCAAGCCCTTCTGATGACA	121
Db	121	GAGTTCAGTTCATATGTGTGAGTGGCTTTTGACCCCTTCCCGCTCAAGCCCTTCTGATGACA	180
OY	1214	AGGCCCTTGGGGTACCAAGCTGC	1234
Db	181	AGGCCCTTGGGGTACCAAGCTGC	201

```

RESULT 15
US-60-452-680-42977
: Sequence 42977, Application US/60452680
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: GROPE, Andrew
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001450
: CURRENT APPLICATION NUMBER: US/60/452,680
: CURRENT FILING DATE: 2003-03-07
: NUMBER OF SEQ ID NOS: 116213
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 42977
:
: LENGTH: 201
:
: TYPE: DNA
: ORGANISM: Homo sapiens
US-60-452-680-42977

```

Query Match	13.3%	Score 200.6;	DB 12;	Length 201;
Best Local Similarity	99.5%;	Pred. No. 5.7e-48;		
Matches 200; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

QY	357	TTACCTGCTGTAAGGCTCACTTTGAATGTGTGCTCCCTCGAAGAAAGCTGGAGATGCC	416
Db	1	CTACCTGCTGTAAGGCTCACTTTGAATGTGTGCTCCCTCGAAGAAAGCTGGAGATGCC	60
QY	417	CCCATCTCTGGTGTGGAGCGTGATGGCGGTCACTATGATGTCGGGGCAACACTGGAGACAA	476
Db	61	CCCATCTCTGGGTTGGAGCGGTGATGGCGGTCACTATGATGTCGGGGCAACACTGGAGACAA	120
QY	477	GAACTCTGTGATGGCATTACTGACAGCCCTTGGAGCTCCGTGCGTGATCAGAGATACATCC	536
Db	121	GAACTCTGTGATGGCATTACTGACAGCCCTTGGAGCTCCGTGCTGATCAGAGATACATCC	180

Tue Jul 1 11:49:15 2003

us-10-014-896-1.rnpn

Page 12

OY 537 CCGAAGACCTTCTTCATTC 557
|||||
DB 181 CCGAAGACCTTCTTCATTC 201

Search completed: July 1, 2003, 09:11:02
Job time : 1251 secs

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	1373.4	91.0	2101	9	AK057131	AK057131 Homo sapi
2	1310.2	86.8	1422	6	AX319670	AX319670 Sequence
3	1310.2	86.8	1779	6	AX480926	AX480926 Sequence
4	673.6	44.6	3071	10	BC025830	BC025830 Mus muscu
5	235.2	15.6	188119	9	AC119673	AC119673 Homo sapi
6	186	12.3	185081	2	AC079937	AC079937 Mus muscu
7	173	11.5	15633	2	AL359702	AL359702 Homo sapi
8	170.4	11.3	213050	1	AL646079	AL646079 Ralstonia
9	131.2	8.7	171999	2	AC121225	AC121225 Rattus no
c	125.6	8.3	156313	2	AL359702	AL359702 Homo sapi
c	11	8.0	134617	2	AC115001	AC115001 Mus muscu
12	110.2	7.3	5664	1	AE005748	AE005748 Caulobact
13	106.2	7.0	86564	1	AE008919	AE008919 Unculture
14	100.2	6.6	171939	2	AC121225	AC121225 Rattus no
15	93.2	6.2	188119	9	AC119673	AC119673 Homo sapi
16	59.2	3.9	38666	8	SPAC24C9	Z88601 S.pombe chr
17	56	3.7	31360	1	SC9C7	AL035161 Streptomy
18	51.4	3.4	1469	10	BC005631	BC005631 Mus muscu
19	51	3.4	51	6	AX157288	AX157288 Sequence
20	50.8	3.4	1221	6	E04019	E04019 DNA sequence
21	50.8	3.4	1304	4	SCAC1	X88564 S.scrofa mR
22	50.8	3.4	1341	4	PIGAC1	D13514 Porcine mRN
23	49.4	3.3	51	6	AX157285	AX157285 Sequence
24	49.4	3.3	51	6	AX157287	AX157287 Sequence
25	49.2	3.3	1224	6	E04020	E04020 cDNA sequen
26	49.2	3.3	1224	6	E07281	E07281 genomic DNA
27	49.2	3.3	1224	6	I18856	I18856 Sequence 3
28	49.2	3.3	1224	6	I32240	I32240 Sequence 3
29	49.2	3.3	1388	9	H0MACY1	D14524 Human N-ac
30	49.2	3.3	1403	9	H0MACY1	D16307 Human mR
31	49.2	3.3	1415	9	H0MACY1	L07548 Human amino
32	49.2	3.3	1416	9	BC014112	BC014112 Homo sapi
33	49.2	3.3	1423	9	BC003033	BC003033 Homo sapi
34	49.2	3.3	1432	9	BC005045	BC005045 Homo sapi
35	48.8	3.2	10861	1	AE005988	AE005988 Caulobact
36	47.8	3.2	51	6	AX157286	AX157286 Sequence
37	47.6	3.2	12249	1	AE002039	AE002039 Deinetococ
38	46.2	3.1	381	6	AX307444	AX307444 Sequence
39	46.2	3.1	7066	1	SC565	AL591823 Sequence
c	45.6	3.0	147215	8	ATF20D10	AL591823 Streptomy
41	45.6	3.0	139789	8	ATCHR189	AL035538 Arabidops
42	44.8	3.0	663	6	AX364389	AL161593 Arabidops
43	44.8	3.0	799	6	AX364030	AX364030 Sequence
44	44.8	3.0	787	6	AX364483	AX364483 Sequence
45	44.8	3.0	1542	8	AY074847	AY074847 Arabidops

AUTHORS

AK057131 2101 bp mRNA linear PRI 01-AUG-2000
Homo sapiens CDNA FLJ32619.1 clone SPLEN2000134, weakly similar
to CARBOXYPEPTIDASE S PRECURSOR (EC 3.4.17.4)..
AK057131
AK057131.1 GI:16552718
oligo capping; fls (full insert sequence).
Homo sapiens spleen CDNA to mRNA, clone_1lb:SPLEN2
clone:SPLEN2000134.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Takawami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R.,

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 188119)
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 REFERENCE 3 (bases 1 to 188119)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphumachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
 and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUN-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 COMMENT On Jul 16, 2002 this sequence version replaced gi:20340494.

----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchgs@u.washington.edu
 Drafting Center: SC

----- Project Information
 Center project name: chr-1
 Center clone name: RP11-212H11 (sc0653)

----- Summary Statistics
 Sequencing vector: Plasmid; 100% of reads
 Chemistry: Dye-terminator ET; 52% of reads
 Chemistry: Dye-terminator Big Dye; 48% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 187572 bases at least Q40
 Consensus quality: 186068 bases at least Q30
 Consensus quality: 188119 bases at least Q20
 Insert size: 188119; sum-of-contigs
 Quality coverage: 9.0x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:
 5': RP11-131E5 (UWGC:sc0638) AL359702
 3': Mapping in progress

----- Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:
 This sequence has been validated by multiple Complete Digest
 fingerprinting. Comparison of the experimentally derived digest
 fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and
 vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp)
 are not resolved in the fingerprint and hence do not appear
 in the table. There are no significant remaining discrepancies
 between the experimental and predicted values. Uniquely ordered
 fragments are separated by dashed lines.

ECORI

HindIII

BglII

SeqDermap FngPrnt SeqDermap FngPrnt SeqDermap FngPrnt

8696	9085	2931	2918	3195	3145
6	<800	6382	6461	2067	2183
2464	2523	512	<800	4754	4905
14377	14333	449	<800	124	<800
7675	7840	1646	1601	1936	1957
4109	3986	13415	13198	3410	3414
1650	1641	1353	1310	9341	9345
1641	1641	9257	9231	4771	4905
802	852	3345	3303	7995	8199
8094	8674	104	<800	2778	2982
3568	3459	9998	9715	19698	19775
10033	9904	2742	2741	354	<800
4361	4300	806	812	6508	6419
259	<800	8007	8479	3029	2982
5309	5305	1307	1310	971	955
4297	4300	1065	1052	7346	7329
3090	3024	2103	2099	3818	3591
6150	6153	1931	1882	10426	10302
2359	2423	1631	1601	8307	8199
9249	9904	74	<800	5652	5672
1675	1641	436	<800	2212	2183
379	<800	6462	6461	4405	4386
592	<800	7817	7715	13198	13391
703	<800	8986	9231	3380	3414
522	<800	7688	7715	2483	2580
2893	3024	271	<800	2597	2777
4534	4468	5092	5027	5669	5672
7872	7840	2669	2741	5005	4905
844	852	8570	8909	5900	5924
1016	1018	2727	2741	3565	3414
1735	1641	3111	3145	14307	14103
3979	3986	734	812	742	749
5380	5305	4248	4170	3148	3145
2672	2672	4484	4405	6733	6678
14285	14333	5369	5318	1704	1668
13063	12763	6250	6461	229	<800
2586	2672	3010	3048	4457	4386


```

-----
10810      10492      5321      5318      898      868
-----
397        <800      6530      6461      2021      1957
-----
476        <800      5041      5027      5046      4905
-----
4635       4600      9981      9715      1246      1315
-----
435         <800      2324      2332      1393      1315
-----
215         <800      1068      1052
-----
326         <800      4868      4769
-----
2417        2423      4780      4769
-----
312         <800      2085      2099
-----
3837        3783      38      <800
-----
3711        3783      2320      2332
-----
310         <800      808      812
-----
1386        1345      4672      4614
-----
753         761
-----
1900        1875
-----
1879        1875
-----
FEATURES
SOURCE      location/Qualifiers
            1. 188119
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="RP11-212H11"
            /clone_11b="RPC1 human BAC library 11"

Query Match      15.6%; Score 235.2; DB 9; Length 188119;
Best Local Similarity 96.8%; Pred. No. 3.1e-56;
Matches 240; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 248 TTCATTAAGCTCTTCTCTACAGTGTGCACACACAGCTTATCCAGCATGAGTCGTGAG 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172804 TGCTTCTAGTCTTCTCTACAGTGTGCACACACAGCTTATCCAGCATGAGTCGTGAG 172745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 308 AGTATAGCCACCTGTTCACTATCCAGGCTGGAGCCCAAGCTTGCAGCCCTAAGTCTGA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172744 AGTATAGCCACCTGTTCACTATCCAGGCTGGAGCCCAAGCTTGCAGCCCTAAGTCTGA 172685
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 368 TGGGCACCTTGTATGTGCTGCTGCCCTGGAAGAGGTGGAGGTGCCCCATTCCTG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172684 TGGGCACCTTGTATGTGCTGCTGCCCTGGAAGAGGTGGAGGTGCCCCATTCCTG 172625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 428 GGTGAGAGCTGATGCGCTCATATGTGTGGGGACACTGAGACACAAGAAGCTGTGGA 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172624 GGTGAGAGCTGATGCGCTCATATGTGTGGGGACACTGAGACACAAGAAGCTGTGGA 172565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 488 TGGCATTA 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172564 TGGTCTGA 172557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AC079937      185081 bp      DNA      linear      HMG 19-SEP-2000
LOCUS
DEFINITION
AC079937      Mus musculus clone RP23-135H10 strain C57Bl6/J, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
AC079937
AC079937.1  GI:10190771

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

HTG; HTGS_PHASE1; HTGS_DRAFT.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 185081)
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, J., Ho, S.-L.,
Idol, J., Lee-Lin, S.-O., Legaspi, R., Lim, M., Maduro, Q. L.,
Maduro, V.B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Snyder, B., Stantipod, S., Summers, T.J., Thomas, J.W.,
Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., and Green, E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 185081)
Green, E.D.
Direct Submission
Submitted (19-SEP-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
-----
Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nih.gov
Project Information
Center project name: agx
Center clone name: 135H10
-----
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174235 bases at least Q40
Consensus quality: 177968 bases at least Q30
Consensus quality: 179643 bases at least Q20
Insert size: 218000; agarose-fp
Insert size: 183381; sum-of-contigs
Quality coverage: 4.04x in Q20 bases; sum-of-contigs
Quality coverage: 4.80x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2260: contig of 2260 bp in length
2261: gap of unknown length
2361: contig of 2850 bp in length
5211: gap of unknown length
5211: gap of unknown length
8723: contig of 3412 bp in length
8723: gap of unknown length
11723: contig of 2901 bp in length
11724: contig of 2901 bp in length
11824: gap of unknown length
11824: gap of unknown length
15978: contig of 4155 bp in length
15978: gap of unknown length
15979: gap of unknown length
16078: gap of unknown length
16079: gap of unknown length
18643: contig of 2565 bp in length
18643: gap of unknown length
18644: gap of unknown length
23298: contig of 4555 bp in length
23298: gap of unknown length
23398: gap of unknown length
30271: contig of 6873 bp in length
30271: gap of unknown length
30371: gap of unknown length
30372: gap of unknown length
30372: contig of 8100 bp in length
38471: gap of unknown length
38472: gap of unknown length
38571: gap of unknown length
38572: contig of 8947 bp in length
47518: gap of unknown length
47519: gap of unknown length
56601: contig of 8983 bp in length
56701: gap of unknown length
56702: gap of unknown length
68548: contig of 11847 bp in length
68549: gap of unknown length
76640: contig of 7992 bp in length

```

```

* 76641 76740: gap of unknown length
* 76741 88724: contig of 11984 bp in length
* 88725 88824: gap of unknown length
* 88825 101313: contig of 12489 bp in length
* 101314 101413: gap of unknown length
* 101414 111047: contig of 9634 bp in length
* 111048 111147: gap of unknown length
* 111148 136566: contig of 25419 bp in length
* 136567 185081: contig of 48415 bp in length.
* 136667 Location/Qualifiers
      1. 185081
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RP23-135H10"
      /clone_id="RP23 mouse BAC library 23"
      1. 2260
misc_feature /note="assembly_fragment"
2361..5210
/note="assembly_fragment"
5311..8722
/note="assembly_fragment"
8823..11723
/note="assembly_fragment"
11824..15978
/note="assembly_fragment"
16079..18643
/note="assembly_fragment"
clone_end:sp6
vector_slide:sp6
misc_feature /note="assembly_fragment"
18744..23298
/note="assembly_fragment"
23399..30271
/note="assembly_fragment"
30372..38471
/note="assembly_fragment"
38572..47518
/note="assembly_fragment"
47619..56601
/note="assembly_fragment"
56702..68548
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
clone_end:r7
vector_slide:right"
68649..76640
/note="assembly_fragment"
76741..88724
/note="assembly_fragment"
88825..101313
/note="assembly_fragment"
101414..111047
/note="assembly_fragment"
111148..136566
/note="assembly_fragment"
136667..185081
/note="assembly_fragment"
BASE COUNT 48829 a 43005 c 42959 g 48571 t 1717 others
ORIGIN

```

```

Query Match 12.3%; Score 186; DB 2; Length 185081;
Best Local Similarity 85.5%; Pred. No. 5.7e-42;
Matches 207; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

```

QY 249 TCATTAAGTCTTCTCTACAGTGTGACACACAGCTTTATCCAGCATTAAGTCTGGAAGA 308
DB 162960 TCCCTCAAGCTTCCCTCAAGTGTGACACACAGCTTCTCAACATGAAGTCTGGCAAA 163019
QY 309 GTATAGCACCTGTGTACTATCAAGGCTGAGACCCAGCTGAGAGCCCTACCTGCGTAT 368
DB 163020 GTATAGCACCTGTGTACTATCAAGGCTGAGACCCAGCTGAGAGCCCTACCTGCGTAT 163079
QY 369 GGCCTCATTTGATGTGTGCTGCCCTGAAAGAGGCTGGAGAGTGGCCCCCATTTCTGTGG 428

```

```

DB 163080 GGCCTCATTTGATGTGTGCTGCCCTGAAAGAGGCTGGAGAGTGGCCCCCATTTCTGTGG 163139
QY 429 GTTGACAGCTGATGAGGCTCATATGATGAGGAGGACACATGACAGCAAGAACTGTGAT 488
DB 163140 CCTGGAACGAAATGCTTATCATATGAGGAGGCTGAGACAAACAAACTGTGAT 163199
QY 489 GG 490
DB 163200 GG 163201

```

```

RESULT 7
AL359702 156313 bp DNA 1linear HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-131E5, *** SEQUENCING IN
DEFINITION
ACCESSION AL359702
VERSION AL359702.4 GI:9798027
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 156313)
REFERENCE
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: cloneenquiries@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:977731.
----- genome center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: D1J1E5
Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 151344 bases at least Q40
Consensus quality: 154167 bases at least Q20
Insert size: 155213; sum-of-contigs
Insert size: 156039; 0.2% error; agarose-fp
Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality
coverage: 4.72x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 12576: contig of 12576 bp in length
* 12577 12676: gap of 100 bp
* 12677 17994: contig of 5318 bp in length
* 17995 18094: gap of 100 bp
* 18095 54844: contig of 36750 bp in length
* 54845 54944: gap of 100 bp
* 54945 70607: contig of 15663 bp in length
* 70608 70707: gap of 100 bp
* 70708 91781: contig of 21074 bp in length
* 91782 91881: gap of 100 bp
* 91882 104123: contig of 12242 bp in length
* 104124 104223: gap of 100 bp
* 104224 106666: contig of 2443 bp in length
* 106667 106767: gap of 100 bp
* 106767 112670: contig of 5904 bp in length

```


* 112671 112770: gap of 100 bp
* 112771 132997: contig of 20227 bp in length
* 132998 133097: gap of 100 bp
* 133098 135495: contig of 2398 bp in length
* 135496 135595: gap of 100 bp
* 135596 153621: contig of 18026 bp in length
* 153622 153721: gap of 100 bp
* 153722 156313: contig of 2592 bp in length.
Location/Qualifiers
1. 156313

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-131E5"
/clone.lib="RP11-131E5"

misc_feature

1. 12576
/note="assembly_fragment:01662
fragment_chain:1"

misc_feature

12677. 17994
/note="assembly_fragment:01305
fragment_chain:1"

misc_feature

18095. 54844
/note="assembly_fragment:00421
fragment_chain:1"

misc_feature

54945. 70607
/note="assembly_fragment:00007
fragment_chain:1"

misc_feature

70708. 91781
/note="assembly_fragment:00483
fragment_chain:1"

misc_feature

91882. 104123
/note="assembly_fragment:00088
fragment_chain:1"

misc_feature

104224. 106666
/note="assembly_fragment:01682
fragment_chain:2"

misc_feature

106767. 112670
/note="assembly_fragment:00618
fragment_chain:2"

misc_feature

112771. 132997
/note="assembly_fragment:00134"
fragment_chain:2"

misc_feature

133098. 135495
/note="assembly_fragment:00993"
fragment_chain:3"

misc_feature

135596. 153621
/note="assembly_fragment:00859
fragment_chain:3"

misc_feature

153722. 156313
/note="assembly_fragment:01584
fragment_chain:3"

BASE COUNT

43117 a 35613 c 34109 g 42363 t 1111 others

ORIGIN

Query Match 11.5%; Score 173; DB 2; Length 156313;
Best Local Similarity 97.2%; Pred. No. 3.3e-38;
Matches 176; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY

1110 CCAAGAGTCTCTAAGTCAAGAGGACATTTGGTGGATTAACAGAGTCCAGTTCATGT 1169

DB

133344 CCTGGAGTCTCTAAGTCAAGAGGACATTTGGTGGATTAACAGAGTCCAGTTCATGT 133403

OY

1170 GTTGAAGTCTCTAAGTCAAGAGGACATTTGGTGGATTAACAGAGTCCAGTTCATGT 1229

DB

133404 GTTGAAGTCTCTAAGTCAAGAGGACATTTGGTGGATTAACAGAGTCCAGTTCATGT 133463

OY

1230 GCTGCTCCGCGACGACCGTACAGTCCGCTTCCTCCGGAAGTCAATTTACTGCCCAAGTTAC 1289

DB

133464 GCTGCTCCGCGACGACCGTACAGTCCGCTTCCTCCGGAAGTCAATTTACTGCCCAAGTTAA 133523

OY

1290 T 1290

DB

133524 T 133524

RESULT 8

AL646079

213050 bp DNA linear

BCT 07-DEC-2001

LOCUS

Ralstonia solanacearum GM11000 megaplasmid, complete sequence;

DEFINITION

segment 4/11.

ACCESSION

AL646079 AL646053

VERSION

AL646079.1 GI:17430956

KEYWORDS

Ralstonia solanacearum.

SOURCE

Ralstonia solanacearum.

ORGANISM

Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

REFERENCE

1 (bases 1 to 213050)

AUTHORS

Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,

TITLE

Arlet,M., Billault,A., Bottier,P., Camus,J.C., Cattolico,L.,

JOURNAL

Chandler,M., Choisme,N., Claudel-Renard,C., Cunne,S., Demange,N.,

REFERENCE

Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schlex,T.,

AUTHORS

Siglier,P., Thebaud,P., Whalen,M., Winkler,P., Levy,M.,

TITLE

Weissenbach,J. and Boucher,C.A.

JOURNAL

Genome sequence of the plant pathogen Ralstonia solanacearum

REFERENCE

Unpublished

AUTHORS

2 (bases 1 to 213050)

TITLE

Boucher,C.A.

JOURNAL

Direct Submission

REFERENCE

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston

AUTHORS

Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie

TITLE

Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,

JOURNAL

BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean

REFERENCE

Dausset-CEPH, 27 rue Juliette Doda, 75010 Paris, France, LMCM CNRS

AUTHORS

118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA

TITLE

URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,

JOURNAL

Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,

REFERENCE

F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire

AUTHORS

INRA, BP27, F31326 Castanet-Tolosan Cedex

TITLE

Christian.Bouchere@toulouse.inra.fr

JOURNAL

http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES

Location/Qualifiers

source

1. 213050

COMMENT

/organism="Ralstonia solanacearum"

FEATURES

/strain="GM11000"

source

/db_xref="taxon:305"

gene

/plasmid="megaplasmid"

CDS

100. 1632
/gene="RSP0487"
/note="RSP0365"
100. 1632
/gene="RSP0487"
/function="macromolecule metabolism; macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD17638.1"
/db_xref="GI:17430957"
/translation="MDAPSGRGAVIRAVGRLLVTVGLAVLAFAFTWRDNR
GLDVPAPLAVGAAALRLGLAIPRTISFTDAQONADQFRLHMLATRYRAH
AAMORPVDGFSILTYWKGSDPSIRPILMAHODVPVAPRGEGMTEPPFAGVYVADG
MWRGAMPDKLILQMAELIASGRPRRTIRFAGADEBVGEGRGARIALLL
KSRGDLAVYIDGGLITGEGVPGTLRPALIGVTEKGLSVALKISATPGHSSMPPA
PGESATAMSAALAKHLDQDLPAIGIRGVREKFTLAPRGRLNRYLSNMLTLEFLV
RKOLASPSNNAVLTQTTALTIVQAGNKDNYLPGRAETAVNRLPLGDSASVIAHVE
QAVSAVPGKHETALPGVSEAPVSPVPSOSYDILGRTVEVPGVYVAGLVGA
TDSRHMGISDHYFRFSPVRAREDIARHGNERISEANVELIRFYRLIHOADA
GA"

gene
complement(1751..3148)
/gene="czcs"
/note="Rsp0488; RS00366"
complement(1751..3148)
/gene="czcs"
/EC number="2.7.3.-"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17639.1"
/db_xref="GI:17430958"
/translation="MNRMRGSLARLVLAFLVYVAGVMTGAGTYIYALNQLRATDD
AMGICVAPLRHCAKDQSDIDIVDARRTVEVSGHGLILRTVIGPKRTLVANPE
GEVNPAAVPPDVLASESALHTWMSKRGISRAIAQARIGNHEDYVTVADGDS
RAALMSYRHRIFLAVIAGVAVALISLIMVSNLTKGARIGROASVPEQLTRTLE
LRGAPRELSDLLAALNMTLRIDGFPARLSQASADLADFPPLANLIGOTEVIAH
RAGEFRVATESLEEARLSRIMEMFLARADARVALSVAPLADAEHIEIVYDFR
EVAADRDITVICGNGRYRADLTLLRAINNLDMALRHPGARDIVANVYPSHT
ATAVRNPGDIPAAISPLIFDFYRSGDPAARANSERSAGLGAIYQTIIMDLHGTAQAS
SPGGLTERLTPPA"
complement(3145..3819)
/gene="czcr"
/note="Rsp0489; RS00367"
complement(3145..3819)
/gene="czcr"
/function="cell processes; protection responses"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17640.1"
/db_xref="GI:17430959"
/translation="MRLLIVEDPEKAGDYLLKGLTESGFVADLARSGPDGIVKFP
YDLIVDLMLGEGMGMQVIRIELRRKTKTPVFLTAPBELSRKLGEELCADHIVKFP
AAEELVAVYHTLRGPMRESFVEIADITDTIKRRTLRGRRTIDLTAAEVALHLIL
ARTEVYLSRSLISSQVDMVNFDSNTNVDAIRRLRAKVDPEHRLIHTLRGMGYV
LDTVAR"
4089..4394
/gene="Rsp0490"
/note="RS00368"
4089..4394
/gene="Rsp0490"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17641.1"
/db_xref="GI:17430960"
/translation="MSVRRNCLIPLAAAAFLACGATTAASVEQALSPQLVHGARDP
YTDGAKGRDPTVTCGGADEATITDASVGPDPYISDGMSSRSAPRQDDPTDGAQA"
4400..5725
/gene="czcc"
/note="Rsp0491; RS00369"

[illegible]

```

Db 532 GAGGGGACTGAGACCGAGCCCTTTCGCGGTGTGAGAGCGCATGTGTGGGCG 591
Oy 457 CGGGGACACTGGAGCAGCAAGACTGTGATGGCATTTACTGAGGCTTGGAGCTCTG 516
Db 592 GCGGGGAGTGGAGGAGCAGCAAGGCAACCTGATGCCAGATGAGAGCCCGGCAATTGCTG 651
Oy 517 CTGATCAGCAAGTACATCCCGCAGATCTTCTTCAATTTCTCTGCGGCAATGATGAGAG 576
Db 652 GCGGGGAGGCGGTTCGATCCCGCCGACATCATCTTCCGCTTCTGTCGCGAGAGAG 711
Oy 577 TCATCAG---GACAGGGGCTCAGAGATCTCAGCCCTGCTACAGTCAGAGGGGCTCCAG 633
Db 712 GTCGCGCGGAGCGCGCGCGCGCGCAGATGTCGCTGCTGTAATTCGCGCGAGCAG 771
Oy 634 CTGAGCTTCAATTTGAGCAGAGGGGGGCTTCAATCTTGGATTTCAATCTTCACTTCAAG 693
Db 772 CTGCGCTTGTGATCAGCAGAGGGGTTGCTATACCGAGGGCCTGGTCCGCGGGCTGAGC 831
Oy 694 AAGCCCATTCGCTTGTATTTGAGTCTCAGAGAGGGGTTCCATGAACTCATGCTCAGATGA 753
Db 832 AAGCGGCGCGGCTGATCGCGGCTACCGAGAGAGGCTTCTGTCGGTGCGATTGAAGTTG 891
Oy 754 AACTGACTTCAGGCGCCTTTAGCTCTTCAAG-----GAGCAACCATTTGGCATC 807
Db 892 TCTGCGAGCGCGCGCTTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATG 951
Oy 808 CTTCAGCTGTGTCAGCGCATTTGAGCAGAGCAACCAATGCTATCATATTTGAGAGCGG 867
Db 952 ATGAGCGCGCTTCAAGCACTGAGACAGAGAGAGTCCCGCGCGCATTCGCGGGGTG 1011
Oy 868 ACACTGTGACTGATTTGAGCAACATGCGCAAAATGATTTCCCTTCCCTGTCATATATATC 927
Db 1012 ACCGCGAGAG---TGTCGAGAGCGCTGCGCGCGCGAGATGCGCGCGCGCGCGCGCGCG 1068
Oy 928 CTGAGCAACCGCATTTGAGCACTTATTAAGCGGTTATGAGAGAAATCCCTTA 987
Db 1069 CTGTCCACACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1128
Oy 988 ACCAATGCAATATATCAGAGACACAGCGCACTACCATATTTCAAAAGAGGGGTCAAGTTC 1047
Db 1129 ACCAAGCGCGGTGTCAGAGACACACAGCGCGCTGACATTCGTCAGCGCGCGCAAGAGAC 1188
Oy 1048 AATGTCAATCCCGCAAGTGGCCAGACCATCAATTCGCGATTCACCCGCGAGAGCA 1107
Db 1189 AAGCTCTGCGCGCGCGCGCGCGCGAGGCAAGGTCGCTGCTGCTGCTGCTGCTGCTGCTG 1248
Oy 1108 GTCCAAAGAGTCTTGAAGTCAAGCAAGCAATTTGCTGATTAACAGAGTCCAGTTCAT 1167
Db 1249 GCTTCGTCCGTGATCGCGCAGTGTGAGAGAGCGGTGCGGTGCGCGGTGCCCAAGGGGAT 1308
Oy 1168 GTGTGAGTGCCTTTGACCCCTC---CCGTCAGCCCTTCTGATGACAGAGGCTTGG 1222
Db 1309 TTCGAGCTGCGCGCGCTCCCGCGGTGAGCGAGCGCGCGCTGCTGCTGCTGCTGCTGCTG 1368
Oy 1223 G----CTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278
Db 1369 GCGTCTTCAAGCTATGCGAGGAGGCTGCGAGGAGTTCGCG---GGAGCGGTG 1425
Oy 1279 GCGCCAGTACTTCTATTTGAGCAACAGACAGCGGATTTCTTAAACCTCAGCACTGGC 1338
Db 1426 GCGCGGCGGTGATGAGTGGGCGCGAGCATTCGCGGCAACATGATGCGGATCAGGACACAC 1485
Oy 1339 ATCTCAGAGTCTTACCCATCTACTATACAGCTGTAAGACTTCAAAAGCATCCATGAGATC 1398
Db 1486 GATATTCGCTCTCTCCCGGTGCGCGCGCGGAGATCTGCGCGCTTCCACGCGCAGC 1545
Oy 1399 AACGAGAAATCTCAG 1414
Db 1546 AACGAGCGCATCTCG 1561

```

LOCUS AC121225 171999 bp DNA linear HTG 24-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-89122, *** SEQUENCING IN PROGRESS
 AC121225
 *** 77 unordered pieces.
 AC121225.3 GI:21909154
 HTG; HTGS_PHASE1.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS
 1 (bases 1 to 171999)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaraltinge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barhale,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowle,S., Blevins,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhaey,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douhalwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabrill,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homs,F., Howard,S., Huber,J., Hulys,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolyet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
 Kratochvil,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,Y., Louiseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Meshwari,M., Mapua,P., Martin,R., Merlindade,A., Martinez,E.,
 Messeny,E., Mashiney,E., McLeod,T.P., Meador,M., Mel,G., Metzger,M.,
 Moser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Nguyen,N., Neal,D., Newton,J., Newton,N., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogund,M., Okunolu,G.,
 Oragunye,N., Oylerdo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojuben,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoshari,N., Sisson,I.,
 Sodergren,E., Sonakhe,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telifrod,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 2 (bases 1 to 171999)
 Direct Submission
 Morley,K.C.
 Direct Submission
 Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 171999)
 Morley,K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 19, 2002 this sequence version replaced gi:21240469.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

1	1095:	contig of 1035 bp in length
2	1096:	gap of unknown length
3	1195:	gap of unknown length
4	1196:	2367: contig of 1172 bp in length
5	2368:	2467: gap of unknown length
6	2468:	4128: contig of 1661 bp in length
7	4129:	4228: gap of unknown length
8	4229:	5398: contig of 1170 bp in length
9	5399:	5498: gap of unknown length
10	5499:	6503: contig of 1005 bp in length
11	6504:	6603: gap of unknown length
12	6604:	7990: contig of 1387 bp in length
13	7991:	8090: gap of unknown length
14	8091:	9719: contig of 1629 bp in length
15	9720:	9819: gap of unknown length
16	9820:	11874: contig of 2055 bp in length
17	11875:	11974: gap of unknown length
18	11975:	13052: contig of 1078 bp in length
19	13053:	13352: gap of unknown length
20	13353:	14280: contig of 1128 bp in length
21	14281:	14380: gap of unknown length
22	14381:	15777: contig of 1397 bp in length
23	15778:	15877: gap of unknown length
24	15878:	16879: contig of 1002 bp in length
25	16880:	16979: gap of unknown length
26	16980:	18537: contig of 1558 bp in length
27	18538:	18637: gap of unknown length
28	18638:	20054: contig of 1417 bp in length
29	20055:	20154: gap of unknown length
30	20155:	21560: contig of 1426 bp in length
31	21581:	21680: gap of unknown length
32	21681:	22789: contig of 1109 bp in length
33	22790:	22889: gap of unknown length
34	22890:	24429: contig of 1540 bp in length
35	24430:	24529: gap of unknown length
36	24530:	25932: contig of 1403 bp in length
37	25933:	26032: gap of unknown length
38	26033:	27782: contig of 1760 bp in length
39	27793:	27893: gap of unknown length
40	27893:	29087: contig of 1195 bp in length
41	29088:	29187: gap of unknown length
42	29188:	30744: contig of 1557 bp in length
43	30745:	30844: gap of unknown length
44	30845:	32726: contig of 1892 bp in length
45	32737:	32836: gap of unknown length
46	32837:	34268: contig of 1432 bp in length
47	34269:	34368: gap of unknown length
48	34369:	36154: contig of 1766 bp in length
49	36155:	36254: gap of unknown length
50	36255:	37586: contig of 1332 bp in length
51	37587:	37686: gap of unknown length
52	37687:	38966: contig of 1310 bp in length
53	38967:	39066: gap of unknown length
54	39097:	40283: contig of 1187 bp in length

RESULT 10

AL359702/c
 LOCUS 156313 bp DNA linear HTG 20-JAN-2001
 DEFINITION Homo sapiens chromosome 1 clone RP11-131E5, *** SEQUENCING IN
 PROGRESS ***, 12 unordered pieces.
 ACCESSION AL359702
 VERSION AL359702.4 GI:9798027
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 156313)
 TITLE Direct Submission
 JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: cloneenquiries@sanger.ac.uk
 On Aug 12, 2000 this sequence version replaced gi:8977731.
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: DAL31E5
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid: 108752; 100% of reads
 Chemistry: dye-terminator Big Dye; 100% of reads
 Consensus quality: 151344 bases at least Q40
 Consensus quality: 153189 bases at least Q30
 Consensus quality: 154167 bases at least Q20
 Insert size: 155213; sum-of-contigs
 Insert size: 156039; 0.2% error; agarose-fp
 Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality
 coverage: 4.72x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 12576: contig of 12576 bp in length
 * 12577 12676: gap of 100 bp
 * 12677 17994: contig of 5318 bp in length
 * 17995 18094: gap of 100 bp
 * 18095 54844: contig of 36750 bp in length
 * 54845 54944: gap of 100 bp
 * 54945 70607: contig of 15663 bp in length
 * 70608 70707: gap of 100 bp
 * 70708 91781: contig of 21074 bp in length
 * 91782 91881: gap of 100 bp
 * 91882 104123: contig of 12242 bp in length
 * 104124 104223: gap of 100 bp
 * 104224 106666: contig of 2443 bp in length
 * 106667 106766: gap of 100 bp
 * 106767 112670: contig of 5904 bp in length
 * 112671 112770: gap of 100 bp
 * 112771 132997: contig of 20227 bp in length
 * 132998 133097: gap of 100 bp
 * 133098 135495: contig of 2398 bp in length
 * 135496 135595: gap of 100 bp
 * 135596 153621: contig of 18026 bp in length
 * 153622 153721: gap of 100 bp
 * 153722 156313: contig of 2592 bp in length.
 Location/Qualifiers
 1. 156313
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"

misc_feature /clone="RP11-131E5"
 1. 12576 /clone="RP11-131E5"
 /note="assembly-fragment:01662
 fragment.chain:1"
 misc_feature 12677..17994
 /note="assembly-fragment:01305
 fragment.chain:1"
 misc_feature 18095..54844
 /note="assembly-fragment:00421
 fragment.chain:1"
 misc_feature 54945..70607
 /note="assembly-fragment:00007
 fragment.chain:1"
 misc_feature 70708..91781
 /note="assembly-fragment:00483
 fragment.chain:1"
 misc_feature 91882..104123
 /note="assembly-fragment:00088
 fragment.chain:1"
 misc_feature 104224..106666
 /note="assembly-fragment:01682
 fragment.chain:2"
 misc_feature 106767..112670
 /note="assembly-fragment:00618
 fragment.chain:2"
 misc_feature 112771..132997
 /note="assembly-fragment:00134"
 133098..135495
 /note="assembly-fragment:00993"
 135596..153621
 /note="assembly-fragment:00859
 fragment.chain:3"
 misc_feature 153722..156313
 /note="assembly-fragment:01584
 fragment.chain:3
 clone.end:77
 vector.side:right"
 BASE COUNT 43117 a 35613 c 34109 g 42363 t 1111 others
 ORIGIN
 Query Match 8.3%; Score 125.6; DB 2; Length 156313;
 Best Local Similarity 97.0%; Pred. No. 1.8e-24;
 Matches 126; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1378 TTCAAAGCAGCCATGAGTCAAGAGAAATCTCAGTCCAGCTATGAGACCAAGTG 1437
 ||| |||||
 Db 148883 TTCCTCAGATTCATGAGTCAAGAGAAATCTCAGTCCAGCTATGAGACCAAGTG 148824
 Oy 1438 AATTCATCTTTGAGTTCAGATGCTGACACAGACGACGAGTTCCTACCTG 1497
 |||||
 Db 148823 AATTCATCTTTGAGTTCAGATGCTGACACAGACGAGGCGAGTTCCTACCTG 148764
 Oy 1498 CACAACTGTGA 1509
 |||||
 Db 148763 CACAACTGTGA 148752
 RESULT 11
 AC115001 134617 bp DNA linear HTG 06-JUN-2002
 LOCUS AC115001
 DEFINITION Mus musculus clone RP24-128M9, WORKING DRAFT SEQUENCE, 15 ordered
 pieces.
 ACCESSION AC115001
 VERSION AC115001.2 GI:21327577
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULITOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 134617)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Mus musculus, clone RP24-128M9

JOURNAL
REFERENCE
AUTHORS

Unpublished

2 (bases 1 to 134617)

Anderson, S., Barna, N., Nussbaum, C., Lander, E., All, A., Allen, N.,
Birkner, B., Linton, L., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, K., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Nobdu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhng, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

JOURNAL
REFERENCE
AUTHORS

Submitted

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 134617)

Anderson, S., Barna, N., Nussbaum, C., Lander, E., All, A., Allen, N.,
Birkner, B., Linton, L., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, K., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Nobdu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhng, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 124383

Center clone name: 128_M.9

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 129731 bases at least Q40

Consensus quality: 131917 bases at least Q30

Consensus quality: 132753 bases at least Q20

Insert size: 134000; agarose-IP

Insert size: 133217; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; agarose-IP
Quality coverage: 5.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1153: contig of 1153 bp in length
1154: 1253: gap of 100 bp
1254: 3428: contig of 2175 bp in length
3429: 3528: gap of 100 bp
3529: 4365: contig of 837 bp in length
4366: 4465: gap of 100 bp
4466: 7977: contig of 3512 bp in length
7978: 8077: gap of 100 bp
8078: 12322: contig of 4245 bp in length
12323: 12422: gap of 100 bp
12423: 14768: contig of 2346 bp in length
14769: 14868: gap of 100 bp
14869: 19859: contig of 4991 bp in length
19860: 19959: gap of 100 bp
19960: 25158: contig of 5199 bp in length
25159: 25258: gap of 100 bp
25259: 33912: contig of 8654 bp in length
33913: 34012: gap of 100 bp
34013: 42066: contig of 8054 bp in length
42067: 42166: gap of 100 bp
42167: 51449: contig of 9283 bp in length
51450: 51549: gap of 100 bp
51550: 62734: contig of 11185 bp in length
62735: 62834: gap of 100 bp
62835: 80340: contig of 17506 bp in length
80341: 80440: gap of 100 bp
80441: 107825: contig of 27385 bp in length
107826: 107925: gap of 100 bp
107926: 134617: contig of 26692 bp in length.

FEATURES

source

1..134617
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-128M9"
/clone="RP24-128M9"
1..1153
/note="assembly_fragment"
1254..3428
/note="assembly_fragment"
3529..4365
/note="assembly_fragment"
4466..7977
/note="assembly_fragment"
8078..12322
/note="assembly_fragment"
12423..14768
/note="assembly_fragment"
14869..19859
/note="assembly_fragment"
19960..25158
/note="assembly_fragment"
25259..33912
/note="assembly_fragment"
34013..42066
/note="assembly_fragment"
42167..51449
/note="assembly_fragment"
51550..62734
/note="assembly_fragment"
62835..80340

gene
CDS
/gene="EBAC000-65D09.2"
2644..2805
/note="EBAC000-65D09.2"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL76347.1"
/db_xref="GI:18645046"
/translation="MQRRAHNSPANPIIRLAPGLLRVDRATIPVSNPKTIDILIQ
PRTLLGSH"
complement(2791..3786)
/gene="crtc"
/note="EBAC000-65D09.3"
complement(2791..3786)
/gene="crtc"
/note="similar to PID:1809122; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="hydroxyneuroporene synthase"
/protein_id="AAL76348.1"
/db_xref="GI:18645047"
/translation="MGSMPKNGPQIEAGKALGFDAVDPGGYRMWYTLDAFSDGKRA
LTVIVGVSVFSEYRARRRRHAGNENCSVALIYGPDRWMSLTREGACGLKGC
DOIVIGSVSESSYRARRRRHAGNENCSVALIYGPDRWMSLTREGACGLKGC
WMPAAPAGRYKTKMKRGIAMEGAAIDCAAGVYLETETLNHMQRSAAEGSGITH
YDNTLNGERRSGLRYPDMSCTELADSRDPISTETLRAQRCLCRYSGQPRYTE
TLSEPTYARSGLDPTGLGPRKVMHESLIDNFTPRWMQMLLPRMRPRPLVRP"
3820..4761
/gene="ispA"
/note="EBAC000-65D09.4"
3820..4761
/gene="ispA"
/note="identified by match to PFAM protein family HM
PF00348"
/codon_start=1
/transl_table=11
/product="geranyltransferase"
/protein_id="AAL76349.1"
/db_xref="GI:18645048"
/translation="MNLVYLFTLQTFPSMEAYAMGESLDCERTLSAIRBAGOG
CPPLAKALDAVFPGGARVPRICIAVALANGNSEPRLSAAALIELHCASLVHD
DLPCFDALERRKPSLHAAGORIALVSGDALIAFAQLAELGSSVTPVMAQYT
AIVAAVGPMGICAGQAWCEPRVEIYHQAQKTAFLFAATCAGAAAGVSEPPR
ELGNSIGEAQVADDDIADATATGKPGICIDVADRPSAVRELGLGQAVKRLERCI
EKGLDSTPACGRALLHDVYQQQSNRFFPGIKHTAA"
4765..5916
/gene="crtc"
/note="EBAC000-65D09.5"
4765..5916
/gene="crtc"
/note="identified by match to PFAM protein family HM
PF00891"
/codon_start=1
/transl_table=11
/product="hydroxyneuroporene methyltransferase"
/protein_id="AAL76350.1"
/db_xref="GI:18645049"
/translation="MSVQGLVSNDRPGYSTWRLRHMYRNNKLIANPRRQMAARTP
LFLRIANRRATQHLHITAGFVYSQLTAAFEVLDLPKLAOGPAQOTDRLISATGPEPA
LFLTLAKAGGLIECEYEGNINWIGELGAAMLANPGIAANVRHQBHLXADLSPTALL
ROREKATLGNWMAVDYKGDISPETSELMATSGOGLADYVLDYVFNKARLVYDA
GGCAFAABAVAPRPNTATVYDLPATYBOAAAHNSNDPSNALOYOGGMPEDPLPE
QADIMSLVRVLHDDKPAQHLLIKAKQALSLNQMLIAIPMAETPSESIGHTYGEF
YLMWGGSRPTSELTSMKTAGFNNVRKASPMPLVAVLIGDKINA"
6037..6984
/gene="bchC"
/note="EBAC000-65D09.6"
6037..6984

gene
CDS
/gene="bchC"
/note="identified by match to PFAM protein family HM
PF0107"
/codon_start=1
/transl_table=11
/product="2-deacetyl-2-hydroxyethyl
bacteriochlorophyllide A dehydrogenase"
/protein_id="AAL76351.1"
/db_xref="GI:18645050"
/translation="MTQVATAVFESPCNLSREVLPCDCAADCLVEIDYSGISTGT
ERLMDGNMHPHPLGNLYPLVGEVGVGHVDSRSKIKAGSVYFVPSRGEFVAGL
RGAARXLYVSARLYQAQPMGSEGVLLIATRVHAKRFPENALPOLYVGYGYS
RIVARIQSLGAEEDRVNQDPARDSGTPTVVDLPKDDTISRRETRICVDGDSYLE
TLFSLSKGQVIVLAGFYHKLQGFDFIAFSEATRLIAAMQEPDELFARSLITDNR
IEVDLITHQYVADAAVADKAFSDPLCKMLIDMRAY"
6981..7976
/gene="bchx"
/note="EBAC000-65D09.7"
6981..7976
/gene="bchx"
/note="identified by match to PFAM protein family HM
PF0142"
/codon_start=1
/transl_table=11
/product="chlorophyllide reductase subunit"
/protein_id="AAL76352.1"
/db_xref="GI:18645051"
/translation="MSPSEAQYVDPGLEASGRADMIASVNEAPKTOIIAIVGKGI
GSEFTLSNLSYMAAQGKRYLLIGDDPSDTSLEFGKACPTIETAAKKAAGDEV
RVEDCQKRDGVYATGELREYRGCGRGITIHGETELGCFHMDVDVYLDPLGD
VYCGGFGPLIADKQKRYIVASNDQSLYANNVCSAVERFRKAGVYAGVATYRK
DVGSGAQNFCAVGIPELASIPANDIRKSNANYEITGPESWGLFAGLASVVAE
APPAPAPALVQGLLEFDGTVGRNVYLQPATLADLDVYKLEQPSLEVYVDV"
7951..9468
/gene="bchY"
/note="EBAC000-65D09.8"
7951..9468
/gene="bchY"
/codon_start=1
/transl_table=11
/product="chlorophyllide reductase subunit"
/protein_id="AAL76353.1"
/db_xref="GI:18645052"
/translation="MRSSTTMSDHSNSPETIPLVADNATDAPGCGSSRTIROAG
QANQETLDVADYVPDPDPOSCPAFSLVGLRMQRTATVLSGACVYGLFT
SHEYGARVGVYPPDSESLVYGLKFEEDIREVHTLADPAKYDAVVNLICVPTSGY
PLDLERSINGVRIIGIDVPGCVTHAARVLAAMLKVARSVSGAPVAPVSLH
SBSKRSIALDEIFPVDAVIDRLRPLGASAGVVPFRERELXALDSLVAMH
SPYSACNRETVAGKPVDSAPVGAEGTEMLTAIGTALGPDADMTYHROTIRSA
EAIANPITARNTLSGYESSELVGRILYESCADLRYGSACPGTKMNHDKREMLAH
GYVYKFRASLEDOTCAVDAFEPDLAIGTPVQAKQKGVPSLYFTNLISARPLMGA
GAGLSQVINTAVNNSRPAANDSFEVSGSDAGVWTEDTNRSGLIKREVA"
9468..10910
/gene="bchZ"
/note="EBAC000-65D09.9"
9468..10910
/gene="bchZ"
/note="identified by match to PFAM protein family HM
PF0148"
/codon_start=1
/transl_table=11
/product="chlorophyllide reductase subunit 2"
/protein_id="AAL76354.1"
/db_xref="GI:18645053"
/translation="MLVPHDRAAGYGSVYFTAVKGLVLIIDPVGCGENLPVSYL
HYVDGPNHELPLVVTGLAEEELGKGTGALRAHOTIDARPSVYVYGSIAEMIG
GYVPATNTKRLPTIIDMQOSDRAVLMWDYGGKRRKAKARAVPLVNIIG
PAGYFNTASDQAEIYRLVYEGKARVNRFPGLCHDDIPLSGDADANCYRREGSK
LCOALKPWQAPFGHSTTKFLRALGEEELGADPEEFTQERHTLKPVMDLWRSYTO
DEFGANFGLAATETTYRGRVHRLDEDELPCAFPHFSRKPQVAPNDETIRQTIATSTP
LILFGSYNERMYLAELGKRSIYIPASLPGTIIIRHTGPPKGYSGATYLGVCNALF
DALFHILPLGTEMDAVEATPARVALAMDADKLTETTLVAKQPIIVRISAAKRIDAA
EHSARHQVDVASKEHVLAAGLEVKTGTAA"


```

Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GIV1
Center clone name: CH230-89c22
-----
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 103536 bases at least Q40
Consensus quality: 116344 bases at least Q30
Consensus quality: 124217 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1095: contig of 1095 bp in length
*
*      1096      1195: gap of unknown length
*      1196      2367: contig of 1172 bp in length
*      2368      2467: gap of unknown length
*      2468      4128: contig of 1661 bp in length
*      4129      4228: gap of unknown length
*      4329      5398: contig of 1170 bp in length
*      5399      5498: gap of unknown length
*      5499      6503: contig of 1005 bp in length
*      6504      6603: gap of unknown length
*      6604      7990: contig of 1387 bp in length
*      7991      8090: gap of unknown length
*      8091      9719: contig of 1629 bp in length
*      9720      9819: gap of unknown length
*      9820      11874: contig of 2055 bp in length
*      11875      11974: gap of unknown length
*      11976      13052: contig of 1078 bp in length
*      13053      13152: gap of unknown length
*      13153      14280: contig of 1128 bp in length
*      14281      14380: gap of unknown length
*      14381      15777: contig of 1397 bp in length
*      15778      15877: gap of unknown length
*      15878      16879: contig of 1002 bp in length
*      16880      16979: gap of unknown length
*      16980      18537: contig of 1558 bp in length
*      18538      18637: gap of unknown length
*      18638      20054: contig of 1417 bp in length
*      20055      21580: gap of unknown length
*      21581      21680: gap of unknown length
*      21681      22789: contig of 1109 bp in length
*      22790      22889: gap of unknown length
*      22890      24429: contig of 1540 bp in length
*      24430      24529: gap of unknown length
*      24530      25932: contig of 1403 bp in length
*      25933      26032: gap of unknown length
*      26033      27792: contig of 1760 bp in length
*      27793      27892: gap of unknown length
*      27893      29087: contig of 1195 bp in length
*      29088      29187: gap of unknown length
*      29188      30744: contig of 1557 bp in length
*      30745      30844: gap of unknown length
*      30845      32736: contig of 1892 bp in length
*      32737      32836: gap of unknown length
*      32837      34268: contig of 1432 bp in length
*      34269      34368: gap of unknown length
*      34369      36154: contig of 1786 bp in length
*      36155      36254: gap of unknown length
*      36255      37586: contig of 1332 bp in length
*      37587      37686: gap of unknown length
*      37687      38996: contig of 1310 bp in length

```

Query Match	Best Local Similarity	Score	100.2%	DB 2:	Length	171999:
Matches	114;	Conservative	0;	Mismatches	23;	Indels
						Gaps
						0;
QY	693	GAAGCCCATGCGCTTATGTCAGTCGAGAGAGGCTTCATGAACTCATGCGCAAGT	752			
DB	159979	GATCTCTCCCAACAGATTTTCAGTCCAGAGAGGCTCCCTTACCTCATGCTGCAAGT	160038			
QY	753	AAACATGACTTTCAGGCACTCTTCAGCTCTCCCAAGAAGAGACAAGCATGGCATCTTGC	812			
DB	160039	AAACATGACTTCAGGCACTCTTCAGCTCTCCCAAGAAGAGACAAGCATGGCATCTTGC	160098			
QY	813	AGCTGCTGTGACGCGAAT	829			
DB	160099	TGCGGCTGTGACGCGGT	160115			

RESULT 15
AC119673
LOCUS
DEFINITION
AC119673 AL365208
ACCESSION
VERSION
AC119673.2 GI:21844627
KEYWORDS
HTG.
SOURCE
ORGANISM
human.
Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188119)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimachak, C., Phelps, K.A., Buckley, D., Kidukawa, M., Raymond, C.
and Haugen, E.D.
Direct Submission
Unpublished
2 (bases 1 to 188119)
Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
Direct Submission
Submitted (30-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 188119)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimachak, C., Phelps, K.A., Buckley, D., Kidukawa, M., Raymond, C.
and Haugen, E.D.
Direct Submission
Submitted (16-JUL-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jul 16, 2002 this sequence version replaced gi:20340494.

Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC

Project Information
Center project name: chr-1
Center clone name: RP11-212H11 (sc0653)

Summary Statistics
Sequencing vector: plasmid; 100% of reads
Chemistry: Dye-terminator ET; 52% of reads
Chemistry: Dye-terminator Big Dye; 48% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187572 bases at least Q40
Consensus quality: 188068 bases at least Q30
Consensus quality: 188119 bases at least Q20
Insert size: 188119; sum-of-ctrls
Quality coverage: 9.0x in Q20 bases; sum-of-ctrls

Overlapping Sequences:
5': RP11-131E5 (UWGC:sc0638) AL359702
3': Mapping in progress

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector. In order to accurately represent the entire circular BAC,
small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI

HindIII

BglII

SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint
8696	9085	2931	2918	3195	3145
6	<800	6382	6461	2067	2183
2464	2523	512	<800	4754	4905
14377	14333	449	<800	124	<800
7675	7840	1646	1601	1936	1957
4109	3986	13415	13198	3410	3414
1650	1641	1353	1310	9341	9345
1641	1641	9257	9231	4771	4905
802	852	3345	3303	7995	8199
8094	8674	104	<800	2778	2982
3568	3459	9998	9715	19698	19775
10033	9904	2742	2741	354	<800
4361	4300	806	812	6508	6419
259	<800	8007	8479	3029	2982
5309	5305	1307	1310	971	955
4297	4300	1065	1052	7346	7329
3090	3024	2103	2099	3818	3691
6150	6153	1931	1882	10426	10302
2359	2423	1631	1601	8307	8199
9249	9904	74	<800	5652	5672
1675	1641	436	<800	2212	2183
379	<800	6462	6461	4405	4386
552	<800	7817	7715	13198	13391
703	<800	8986	9231	3380	3414
522	<800	7688	7715	2483	2580
2993	3024	271	<800	2597	2777
4534	4468	5092	5027	5669	5672
7872	7840	2669	2741	5005	4905

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 02:38:10 ; Search time 389 Seconds
(without alignments)
8735.901 Million cell updates/sec

Title: US-10-014-896-1

Perfect score: 1509
Sequence: 1 atggtcagcggtggttg.....ctcactgcacaactgtga 1509

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N Geneseg-101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1310.2	86.8	1421	24	AA597192 Human metalloprotease
2	496.4	32.9	500	21	AAC74554 Human ORFX ORF109
3	267	17.7	319	20	AAK40258 Human secreted pro
4	251	16.6	348	24	ABN24990 Human ORFX polyunc
5	245	16.2	362	22	AA184929 Human polynucleoti
6	51	3.4	51	22	AA173675 Human silent SNP c
7	50.8	3.4	1221	14	AA033105 Pig coding sequenc
8	49.4	3.3	51	22	AA173672 Human silent SNP c
9	49.4	3.3	51	22	AA173674 Human silent SNP c

10	49.2	3.3	1224	14	AA033106 Human coding seque
11	49.2	3.3	1415	24	ABK83807 Human cDNA differe
12	47.8	3.2	51	22	AA173673 Human silent SNP c
13	46.2	3.1	381	24	ABN75263 Human ORF210 cDNA,
14	46	3.0	1445	24	ABL60776 cDNA encoding an e
15	44.8	3.0	663	24	ABK30982 Plant dwarfing/stu
16	44.8	3.0	787	24	ABK30623 Plant dwarfing/stu
17	44.8	3.0	799	24	ABK30776 Arabidopsis thalia
18	44.8	3.0	1493	21	AA033510 Arabidopsis thalia
19	44.8	3.0	1494	21	AA050085 Arabidopsis thalia
20	44.8	3.0	1612	24	ABK30880 Plant dwarfing/stu
21	42.2	2.8	767	21	AA033825 Arabidopsis thalia
22	42	2.8	495	24	AA033895 Human carboxypepti
23	42	2.8	536	24	AA033904 Human carboxypepti
24	42	2.8	613	24	AA033890 Human carboxypepti
25	42	2.8	623	24	AA033886 Human carboxypepti
26	42	2.8	640	24	AA033889 Human carboxypepti
27	42	2.8	665	24	AA033893 Human carboxypepti
28	42	2.8	754	22	AA068896 Human cDNA clone
29	42	2.8	1185	21	AA077081 Human ORFX ORF2636
30	42	2.8	1428	22	AA075199 Nucleotide sequenc
31	42	2.8	1428	24	AA033906 Human carboxypepti
32	42	2.8	1428	24	AA033906 Human metalloprote
33	42	2.8	2643	22	AAH14599 Human cDNA sequenc
34	42	2.8	2643	22	AAH33907 Human carboxypepti
35	42	2.8	2710	22	AAI59038 Human polynucleoti
36	41.6	2.8	1988	22	AAI59591 Human expressed po
37	41.6	2.8	1988	22	AA092224 Novel cDNA encodin
38	41.6	2.8	1997	21	AA078161 Human cancer assoc
39	40.6	2.7	558	24	AA033892 Human diagnostic
40	40.4	2.7	1879	22	AA030319 Nucleotide sequenc
41	40.4	2.7	2221	22	AAH75171 Human polynucleoti
42	40.4	2.7	2657	22	AAI60824 Human expressed po
43	40.4	2.7	2659	22	AAI99566 Human carboxypepti
44	39.8	2.6	594	24	AA033887 Human carboxypepti
45	39.8	2.6	14510	24	ABN59554 Novel human coding

ALIGNMENTS

RESULT 1	AA597192	standard; cDNA, 1422 BP.
XX	AA597192:	
XX	26-FEB-2002 (first entry)	
DE	Human metalloprotease partial DNA sequence #21.	
XX	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;	
XX	vasotrophic; antiinflammatory; analgesic; endocrine; nootropic; tranquiliser;	
KW	hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;	
KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;	
KW	metalloprotease; serine protease; cancer; haematopoietic; breast; colon;	
KW	lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;	
KW	immune-related disease; cardiovascular disease; neuronal disease;	
KW	migraine; sexual dysfunction; mood disorder; attention disorder;	
KW	cognition disorder; hypotension; hypertension; psychotic disorder;	
KW	dyskinesia; metabolic disorder; inflammatory disorder; ss.	
OS	Homo sapiens.	
XX	WO200183782-A2.	
PN	08-NOV-2001.	
XX	04-MAY-2001; 2001WO-0514431.	
XX	04-MAY-2000; 2000US-201879P.	
PR	(SUGC-) SUGEN INC.	

XX 08-FEB-2001 (first entry)
 DT Human ORFX ORF109 polynucleotide sequence SEQ ID NO:217.
 XX
 DE Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 XX vlnerrary; antiparasitic; antiparkinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antineuritic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN K0200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PE 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 PI Shimketa RA, Leach M;
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB40345.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 546; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vlnerrary;
 CC antiparasitic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antineuritic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 500 BP; 114 A; 131 C; 136 G; 117 T; 2 other;

Query Match 32.9%; Score 496.4; DB 21; Length 500;
 Best Local Similarity 99.8%; Pred. No. 6.7e-148;
 Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 141 ACGGCTCGCATGAAAGAGGCGCTGAAAGTGCATTCAGATTCACAGTACTTTAG 200
 DB 3 ACGGCTCGCATGAAAGAGGCGCTGAAAGTGCATTCAGATTCACAGTACTTTAG 62
 OY 201 CTCTAGAGAGTCCAACTACTACAGCCCTGCTGAGTTCGGAATAATCAATCAAGTCTT 260
 DB 63 CTCTAGAGAGTCCAACTACTACAGCCCTGCTGAGTTCGGAATAATCAATCAAGTCTT 122
 OY 261 TCCCTAGAGTGCACACACAGCTTATCCAGATGAATGCTGGAAGATATAGCACT 320
 DB 123 TCCCTAGAGTGCACACACAGCTTATCCAGATGAATGCTGGAAGATATAGCACT 182
 OY 321 GTTCACTATCCAAAGCTCGGACCCAGCTTGACGCCCTACCTGCTGATGCTCACTTGA 380
 DB 183 GTTCACTATCCAAAGCTCGGACCCAGCTTGACGCCCTACCTGCTGATGCTCACTTGA 242
 OY 381 TGTGTCCTGCTGCTGCTGAAAGAGGCTGGAGAGTGCCTATTCCTGCTGGAAGCTTGA 440
 DB 243 TGTGTCCTGCTGCTGCTGAAAGAGGCTGGAGAGTGCCTATTCCTGCTGGAAGCTTGA 302
 OY 441 TGGCGTCACTATGCTGCGGCGCACACTGAGACGACAAAGACTCTGATGGCATTTACTGA 500
 DB 303 TGGCGTCACTATGCTGCGGCGCACACTGAGACGACAAAGACTCTGATGGCATTTACTGA 362
 OY 501 GGCTTGGAGCTCTGCTGATCAGAAATACATCCCGGAGATCTTCTTCTTCTTCT 560
 DB 363 GGCTTGGAGCTCTGCTGATCAGAAATACATCCCGGAGATCTTCTTCTTCTTCTTCT 422
 OY 561 GGGCCATGATGAGAGTCAATCAGGGGACAGGAGCTCAGAGATCTCAGCCCTGCTACAG 620
 DB 423 GGGCCATGATGAGAGTCAATCAGGGGACAGGAGCTCAGAGATCTCAGCCCTGCTACAG 482
 OY 621 AAGGGCGTCCAGCTAGC 638
 DB 483 AAGGGCGTCCAGCTAGC 500

RESULT 3
 AAX40258
 ID AAX40258 standard; cDNA; 319 BP.
 XX
 AC AAX40258;
 XX
 XX 16-JUN-1999 (first entry)
 DT
 DE Human secreted protein 5' EST SEQ ID NO:45.
 XX
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haemopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
 XX
 OS Homo sapiens.
 XX
 PN W09906439-A2.
 XX
 PD 11-FEB-1999.
 XX
 PE 31-JUL-1998; 98WO-IB01233.
 XX
 PR 01-AUG-1997; 97US-0904468.
 XX
 PA (GENST) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 DR WPI: 1999-153700/13.
 DR P-PSDB; AAY11540.
 XX
 PT New nucleic acids encoding human secreted proteins - obtained from

CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.

SO Sequence 51 BP; 12 A; 12 C; 18 G; 9 T; 0 other;

Query Match 3.4%; Score 51; DB 22; Length 51;

Best Local Similarity 100.0%; Pred. No. 4.7e-06;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 GGACGGTGGATGGCTCATCTATGTCGGGACACATGAGCAAGAACTC 482

DB 1 GGACGGTGGATGGCTCATCTATGTCGGGACACATGAGCAAGAACTC 51

RESULT 7

AAQ33105 ID AAQ33105 standard; CDNA to mRNA; 1221 BP.

AC AAQ33105;

DT 06-MAY-1993 (first entry)

DE Pig coding sequence for aminoacylase I.

KW AAm; production; yield; recombinant; ss.

XX Sus scrofa.

OS Key Location/Qualifiers

FT CDS 1..1221

FT /*tag- a

XX JP04330279-A.

XX 18-NOV-1992.

XX 22-MAR-1991; 91JP-0081136.

XX 22-MAR-1991; 91JP-0081136.

XX (TAKI) TAKARA SHUZO CO LTD.

XX WPI; 1993-003494/01.

XX P-PSDB; AAR30458.

XX Polypeptide with aminoacylase (I) activity - has specific base

XX sequence and is used for preparing probe, primer and antibody

XX Claim 2; Page 9; 14pp; Japanese.

XX Pig aminoacylase I cDNA is cloned from pig kidney poly (A) mRNA

XX giving clones lambda pKMA-1 to -10. The longest insert fragment in

XX lambda pKMA-10 (1.4kb) was isolated and sequenced. The sequence

XX may be used to transform Saccharomyces cerevisiae to produce

XX aminoacylase I recombinantly.

XX See also AAQ33106.

XX Sequence 1221 BP; 241 A; 344 C; 381 G; 255 T; 0 other;

Query Match 3.4%; Score 50.8; DB 14; Length 1221;

Best Local Similarity .51.9%; Pred. No. 4e-05;

Matches 140; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

QY 318 CCGTTCACATATCCAGGCTGGAACCCACCTTGACAGCCCTTACCTGATGGCTCACTT 377

DB 183 CGTCTGACCTGCGCGGACCAACCCACACCTCTCTCTCACTTGTCTCACTCCACAC 242

QY 378 TGATGTGTGCTGCTCCCTGTAAGAGGCTGGAGGTGCTCCCTCACTTCTGGGTT--GGA 434

DB 243 AGATGTGTCCTGCTCTTCAAGGAGCATTTGAGTATGACCCCTTTGAGGGCTTCAMGA 302

QY 435 GCGGATGGGCTATCTATGTCGGGACACTGACGACAGAACTGTGATGGCAAT 494

DB 303 TGCAGATGGCTATATCTATGTCGGGACGCGCCACAGAGCAAGATGTCGATCACTA 362

QY 495 ACTGACGAGCTTGGAGCTCTGCTATCAGAGATGATCCCGAAGATCTTTTAT 554

DB 363 CTTGAGGCTGTAGAGGCTGTAAGGTTGAGGCGCCACATTTCCAGAACATCCACAT 422

QY 555 TTCTGAGCCATGATGAGAGTCAATCAGG 584

DB 423 GACCTTTGTGTCAGATGAGAGGTTGAGG 452

RESULT 8

AAI73672 ID AAI73672 standard; DNA; 51 BP.

AC AAI73672;

DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:613.

KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;

KW protein therapy; vaccine; probe; diagnostic assay; detection;

KW quantitation; restorative therapy; polymorphic; ds.

XX Homo sapiens.

XX WO200140521-A2.

XX 07-JUN-2001.

XX 30-NOV-2000; 2000WO-US32758.

XX 30-NOV-1999; 99US-0168138.

XX 29-NOV-2000; 2000US-0726173.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI; 2001-356160/37.

XX Polymorphic nucleic acid sequences, useful in genetic testing and

XX therapy -

XX Claim 1; Page 242; 2653pp; English.

XX AA173060 to AA179867 represent isolated human polymorphic polynucleotide

XX sequences (I), which contain single nucleotide polymorphisms (SNPs).

XX AA173114 to AA173329 represent peptides related to human polymorphic

XX polynucleotide sequences. The sequences can be used in gene and protein

XX therapy, and in vaccine production. (I) and the polypeptides encoded by

XX them may be used in the prevention, diagnosis and treatment of diseases

XX associated with inappropriate expression of polymorphic polypeptides.

XX For example, (I) may be used to treat disorders by rectifying mutations

XX or deletions in a patient's genome that affect the activity of

XX polypeptides by expressing inactive proteins or to supplement the

XX patient's own production of polypeptide. Additionally, (I) and its

XX complementary sequences may also be used as DNA probes in diagnostic

XX assays to detect and quantitate the presence of similar nucleic acids

XX in samples, and therefore which patients may be in need of restorative

XX therapy. The polypeptides encoded by (I) may be used as antigens in the

XX production of antibodies specific for polymorphic polypeptides. The

XX antibodies may also be used to down regulate expression and activity.

XX The antibodies may also be used as diagnostic agents for detecting the

XX presence of polymorphic polypeptides in samples.

XX Sequence 51 BP; 8 A; 9 C; 19 G; 15 T; 0 other;

Query Match 3.3%; Score 49.4; DB 22; Length 51;
Best Local Similarity 98.0%; Pred. No. 1.5e-05;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

420 ATTCCTGGGTTGAGCGGTGATGCGTATCTATGTCGGGACACTGGA 470
1 ATTCCTGGGTTGAGCGGTGATGCGTATCTATGTCGGGACACTGGA 51

Db

RESULT 9
AAI73674
ID AAI73674 standard; DNA; 51 BP.
XX AAI73674;
XX
XX 09-NOV-2001 (first entry)
XX
XX Human silent SNP containing nucleic acid SEQ.615.
XX
XX Human: single nucleotide polymorphism; SNP; genome; gene therapy;
XX protein therapy; vaccine; probe; diagnostic assay; detection;
XX quantitation; restorative therapy; polymorphic; ds.
XX
XX Homo sapiens.
XX
XX MO200140521-A2.
XX
XX 07-JUN-2001.
XX
XX 30-NOV-2000; 2000WO-US32758.
XX
XX 30-NOV-1999; 99US-0168138.
XX PR 29-NOV-2000; 2000US-0726173.
XX
XX (CURA-) CTRAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2001-356160/37.
XX
XX Polymorphic nucleic acid sequences, useful in genetic testing and
XX therapy -
XX
XX
XX Claim 1; Page 242; 2653pp; English.
XX
XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
XX sequences (I), which contain single nucleotide polymorphisms (SNPs).
XX AAI53114 to AAI53329 represent peptides related to human polymorphic
XX polynucleotide sequences. The sequences can be used in gene and protein
XX therapy, and in vaccine production. (I) and the polypeptides encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate expression of polymorphic polypeptides.
XX For example, (I) may be used to treat disorders by rectifying mutations
XX or deletions in a patient's genome that affect the activity of
XX polypeptides by expressing inactive proteins or to supplement the
XX patients own production of polypeptide. Additionally, (I) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acids
XX in samples, and therefore which patients may be in need of restorative
XX therapy. The polypeptides encoded by (I) may be used as antigens in the
XX production of antibodies specific for polymorphic polypeptides. The
XX antibodies may also be used to down regulate expression and activity.
XX The antibodies may also be used as diagnostic agents for detecting the
XX presence of polymorphic polypeptides in samples.
XX
XX
XX Sequence 51 BP; 12 A; 11 C; 18 G; 10 T; 0 other;

Query Match 3.3%; Score 49.4; DB 22; Length 51;
Best Local Similarity 98.0%; Pred. No. 1.5e-05;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

432 GGAGGTATGCGCTCATCTATGTCGGGACACTGACGACAGACTC 482
1 GGAGGTATGCGCTCATCTATGTCGGGACACTGACGACAGACTC 482

Db 1 GGAGGTATGCGCTCATCTATGTCGGGACACTGACGACAGACTC 51

RESULT 10
AAQ33106
ID AAQ33106 standard; cDNA to mRNA; 1224 BP.
XX AAQ33106;
XX
XX 06-MAY-1993 (first entry)
XX
XX Human coding sequence for aminoacylase I.
XX
XX AAm; production; yield; recombinant; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX FT 1..1224
XX CDS /*tag= a
XX
XX JP04330279-A.
XX
XX 18-NOV-1992.
XX
XX 22-MAR-1991; 91JP-0081136.
XX
XX 22-MAR-1991; 91JP-0081136.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX WPI; 1993-003494/01.
XX DR P-PSDB; AAR30459.
XX
XX Polypeptide with aminoacylase (I) activity - has specific base
XX sequence and is used for preparing probe, primer and antibody
XX
XX
XX Claim 2; Page 11; 14pp; Japanese.
XX
XX Human aminoacylase I cDNA was determined. The sequence
XX may be used to transform Saccharomyces cerevisiae to produce
XX aminoacylase I recombinantly.
XX See also AAQ33105.
XX
XX
XX Sequence 1224 BP; 250 A; 353 C; 368 G; 253 T; 0 other;

Query Match 3.3%; Score 49.2; DB 14; Length 1224;
Best Local Similarity 51.9%; Pred. No. 0.00013;
Matches 136; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

318 CCGTTCACATATCCAGAGCTGCGAAGCCAGCTTGACAGCCCTACCTGATGGCTTACTT 377
183 CCGTTCACATATCCAGAGCTGCGAAGCCAGCTTGATCTCTCTCAACTCCACAC 242
378 TGATGTCCTGCTGCCCTGAGAGAGCTGGAGTGGCCCAATCTCTGGGTT---GGA 434
243 GGATGTCCTGCTGCTCTCTCAAGAGACATTTGATACAGACCCTTTGAGGCTTCAAGA 302
435 GCGTATGAGGCTCATATGTCGGGACACTGAGACAGACAACTCTGTATGGCAAT 494
303 TTCTGAGGCTACATCTATGCTCCAGAGGCTGCCAGACATGAGTGGTCAATCCAGTA 362
495 ACTGACAGCTTGGAGCTCTGCTGATCAGAGAGTACATCCCGAAGATCTTTCAT 554
363 CTTGGAAGCTGTGAGGAGCTGAGAGGCGCACCGGTTCCCGAAGCATCCACAT 422
555 TTCTGAGGCTATGAGAG 576
423 GACCTTGTGCTATGAGAG 444

RESULT 11
ABK83807

ID ABK83807 standard; cDNA, 1415 BP.
 AC ABR83807;
 XX
 XX
 DT 14-APR-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #378.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 XX WO200228999-A2.
 PN
 PD 11-APR-2002.
 XX
 XX 03-OCT-2001; 2001WO-US30821.
 PF
 XX 03-OCT-2000; 2000US-237189P.
 PR
 PA (GENE-) GENE LOGIC INC.
 XX
 XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 PT WPI; 2002-435328/46.
 PT
 PT Detecting granulocyte activation by detecting differential expression
 of genes associated with granulocyte activation, which serves as
 diagnostic markers that is useful for monitoring disease states and
 drug toxicity
 XX
 XX
 PS Claim 1: SEQ ID NO. 378; 114pp; English.
 XX
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs: (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part

CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1415 BP; 297 A; 416 C; 416 G; 286 T; 0 other;
	Query Match 3.3%; Score 49.2; DB 24; Length 1415;
	Best Local Similarity 51.9%; Pred. No. 0.00014;
	Matches 136; Conservative 0; Mismatches 123; Indels 3; Gaps 1
OY	318 CCTGTCACTATCCAGGCTCGAGCACCAGCTTGACGCCTAACCTGATGGTCACTT 377
Dd	
OY	244 CGGTGTGACTTCGCCAGGACCAACCTTACATCTCTCCATCTTTGCTCACTCCAC 302
Dd	
OY	378 TGATGTGTGCTGCTCCCCGTGAAGAAGGCTGGAGGTGCCCCCATCTCTGTGGTT--GGA 434
Dd	
OY	304 GGATGTGTGTGCTCTGTTCTTAAGAAACATTGGAGTAGCACACCCCTTGAAGCCTTCAAGA 363
Dd	
OY	435 GCGTGAAGGCTCATCTATGTGTGGGGGCACACTGGACGACAAGAATCTGTGATGGCATT 494
Dd	
OY	364 TTCTGAGGAGCTTACATCTATGCCAGGGGTGCCAGGACATGAAGTCCGTACATCCAGTA 423
Dd	
OY	495 ACTGCAGGCTTTGGAGCTCTGCTGATCAGAAATACATCCCAGAGATCTTCTTCAT 554
Dd	
OY	424 CTGGAAGCTGTGAGGAGCTGAAGGTGAGGGCCACCGGTTCCCAAGAACATCCACAT 483
Dd	
OY	555 TTCTCTGGGCATGATGAGAG 576
Dd	
OY	484 GACCTTTGTGCTGTATGAGAG 505
Dd	
RESULT 12	
AAI73673	
ID	AAI73673 standard; DNA; 51 BP.
AC	
XX	AAI73673;
XX	
DE	09-NOV-2001 (first entry)
XX	
DE	Human silent SNP containing nucleic acid seq:614.
XX	
KW	Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KM	protein therapy; vaccine; probe; diagnostic assay; detection;
KW	quantitation; restorative therapy; polymorphic; ds.
XX	
OS	Homo sapiens.
XX	
PX	WO200140521-A2.
PN	
PD	07-JUN-2001.
XX	
PF	30-NOV-2000; 2000MO-US32758.
XX	
PR	30-NOV-1999; 99US-0168138.
XX	
PR	29-NOV-2000; 2000DS-0726173.
XX	
FA	(CDRA-) CDRA GEN CORP.
XX	
PI	Shinkets RA, Leach M;
XX	
DR	WPI; 2001-356160/37.
XX	
PT	Polymorphic nucleic acid sequences, useful in genetic testing and
FT	therapy -
XX	
PS	Claim 1; Page 242; 2653pp; English.
XX	
CC	AAI73660 to AAI79867 represent isolated human polymorphic polynucleotide
CC	sequences (1), which contain single nucleotide polymorphisms (SNPs).
CC	AAI73660 to AAI79867 represent peptides related to human polymorphic
CC	polynucleotide sequences. The sequences can be used in gene and protein
CC	therapy, and in vaccine production. (1) and the polypeptides encoded by
CC	them may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate expression of polymorphic polypeptides.
CC For example, (1) may be used to treat disorders by rectifying mutations
CC or deletions in a patient's genome that affect the activity of
CC polypeptides in a patient's genome that affect the activity of
CC patients own production of polypeptide. Additionally, (1) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantify the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (1) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
CC
XX
SQ Sequence 51 BP; 9 A; 9 C; 18 G; 15 T; 0 other;
Query Match 3.2%; Score 47.8; DB 22; Length 51;
Best Local Similarity 96.1%; Pred. No. 5e-05;
Matches 49; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 420 ATTCTCTGGGTGGAGCGGTATGCGCCTATGTCGCGGCGACACTGGA 470
Db 1 ATTCTCTGGGTGGAGCGGTATGCGCCTATGTCGCGGCGACACTGGA 51
RESULT 13
ABN75263
ID ABN75263 standard; CDNA; 381 BP.
XX
AC ABN75263;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF210 CDNA, SEQ ID NO:419.
XX
XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
XX disease monitoring; cytokine; cell proliferation; cell differentiation;
XX immune modulation; haematopoiesis regulation; tissue growth;
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; tumour inhibition; bodily characteristics; fertility;
XX behaviour; cancer; proliferative disorder; neurological disorder;
XX cardiovascular disease; immune system disorder; organ transplantation;
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
XX hypothyroidism; cholesterol ester storage disease; infection; vulnerrary;
XX vasotrophic; antipsoriatic; antidiabetic; cystostatic; nootropic;
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
XX cardiact; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
XX dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200130366-A2.
XX
XX 29-NOV-2001.
XX
PD 24-MAY-2001; 2001WO-US17076.
XX
PF 24-MAY-2000; 2000US-206690P.
XX
PR (CURA-) CURAGEN CORP.
XX
PA Leach MD, Shinkets RA;
XX
PI WPI; 2002-106200/14.
XX
DR P-PSDB; ABR31237.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and disorders related to organ
XX transplantation
XX
PS Claim 1; Page 373; 2508pp; English.

CC Sequences ABP31028-ABP35561 represent 4334 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antineoplastic activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
SQ Sequence 381 BP; 111 A; 93 C; 92 G; 82 T; 3 other;
Query Match 3.1%; Score 46.2; DB 24; Length 381;
Best Local Similarity 94.1%; Pred. No. 0.00057;
Matches 48; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1040 TCAGTTCATATGTCATCCGCCAGTGGCCGAGGCCACAGTCACTTCGGA 1090
Db 331 TCATGTCATATGTCATCCGCCAGTGGCCGAGGCCACAGTCACTTCGGA 381
RESULT 14
ABL60776
ID ABL60776 standard; CDNA; 1445 BP.
XX
AC ABL60776;
XX
XX 10-SEP-2002 (first entry)
XX
DE cDNA encoding an enzyme similar to human aminoacylase-1 (ACY-1).
XX
XX Aminoacylase-1; ACY-1; metalloprotein; cytosolic enzyme; human; gene;
XX cystostatic; therapeutic; cancer therapy; ss.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX 1..116 /tag= a
XX FT 5'UTR /tag= b
XX FT 117..1238 /tag= b
XX FT CDS 1239..1445 /tag= c
XX FT 3'UTR
XX FT US6387661-B1.
XX PN 14-MAY-2002.
XX
XX PD

PF 23-MAR-2001; 2001US-0814951.
XX
PR 23-MAR-2001; 2001US-0814951.
XX
PA (PEKE) PE CORP NY.
XX
PI Shao W, Yan C, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-478443/51.
XX
DR P-PSDB; ABB08102.
XX
PT Isolated nucleic acid molecules encoding enzymes similar to human
XX
PT aminocyclase-1, useful as a drug target and diagnostic marker for
XX
PT cancers e.g. T cell leukemias and ovary, brain or lung cancers -
XX
PS Claim 5; Fig 1A; 43pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule encoding
XX
XX enzymes similar to human aminocyclase-1 (ACY-1) (EC 3.5.1.14) (a
XX
XX metalloprotein cytosolic enzyme). The ACY-1 similar polynucleotide and
XX
XX encoded peptide sequences can be used as models for the development of
XX
XX human therapeutic targets, aid in the identification of therapeutic
XX
XX proteins, and serve as targets for the development of human therapeutic
XX
XX agents that modulate enzyme activity in cells and tissues that express
XX
XX the enzyme. ACY-1 has been found to be expressed in humans in the
XX
XX placenta, T cells from T cell leukemia, ovary, brain, lung and leukocyte,
XX
XX and therefore may be a drug target for cancer therapy and act as a
XX
XX diagnostic marker for these cancers. The present sequence represents a
XX
XX cDNA encoding an enzyme similar to human aminocyclase-1 (ACY-1).
SQ
Sequence 1445 BP; 375 A; 410 C; 391 G; 269 T; 0 other;
Query Match 3.0%; Score 46; DB 24; Length 1445;
Best Local Similarity 51.6%; Pred. No. 0.0015;
Matches 131; Conservative 0; Mismatches 120; Indels 3; Gaps 1;
QY 326 CTATCCAGGCTGGACCCAGCTGAGCCCTACCTGCGATGCGCTGATGATGG 385
DB 202 CTGACTATGGACCAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 261
QY 386 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 442
DB 282 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 321
QY 443 GCGTCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
DB 322 GCTACATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 381
QY 503 CTTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
DB 382 CTGTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 441
QY 563 GCCATGATGAGGAG 576
DB 442 TGCCTGATGAGGAG 455
RESULT 15
ABK30982
ID ABK30982 standard; cDNA; 663 BP.
XX
XX ABK30982;
XX
XX 23-APR-2002 (first entry)
XX
XX Plant dwarfing/stunting related cDNA seq ID 396.
XX
XX Plant; ss; dwarfism; stunting; EST; expressed sequence tag;
XX
XX Transgenic Plant; Plant metabolism.
XX
XX Planta.
XX
XX WO200208410-A2.

XX
PD 31-JAN-2002.
XX
XX 20-JUL-2001; 2001WO-US23120.
XX
XX 20-JUL-2000; 2000US-219809P.
XX
XX 20-JUL-2000; 2000US-219810P.
XX
XX (DOMC) DOM CHEM CO.
XX
XX (REDD) REDDY S A.
XX
XX (LARR) LARRINUA M I.
XX
XX (RUEG) RUEGGER M.
XX
XX (WEG) WEGGLARZ T.
XX
XX (BLAK) BLAKESLEE B.
XX
XX (ORIE) ORIEDO V B J.
XX
XX (SAVI) SAVICKAS J P.
XX
XX (MCCR) MCCREY A D.
XX
XX (MILL) MILLER A B.
XX
XX (GACH) GACHOTTE D.
XX
XX (GROS) GROSLEY R.
XX
XX (PELL) PELL R.
XX
XX Reddy SA, Larrinua MI, Ruegger M, Weglarz T, Blakeslee B;
XX
XX Oriedo VB, Savickas JP, McCreary AD, Miller AB, Pogue PG;
XX
XX Della-Cioppa RG, Wolfe MG, Zheng W, Gachotte D, Grosley R, Pell R;
XX
XX WPI; 2002-164823/21.
XX
XX Polynucleotide and amino acid sequences identified in one or more
XX
XX metabolic pathways that lead to dwarfism and stunting in plants, useful
XX
XX in agriculture to create dwarf varieties of any plant species -
XX
XX
XX Disclosure: Fig 9; 717pp; English.
XX
XX The invention relates to polynucleotide and amino acid sequences
XX
XX identified in one or more metabolic pathways that lead to dwarfism and
XX
XX stunting in plants. Also included are vectors comprising the
XX
XX polynucleotides, transgenic plants (including the seed and leaf)
XX
XX transformed with the polynucleotides or vectors, a process for altering
XX
XX the metabolism of a plant comprising providing the above vector and a
XX
XX plant, and transforming the plant with the vector under conditions such
XX
XX that the metabolism of the plant is altered by expression of the isolated
XX
XX nucleic acid from the vector, e.g. such that a stunting phenotype in an
XX
XX industrial plant is produced and a process for the characterization of
XX
XX fractionated biological samples, comprising (a) providing one or more
XX
XX apparatus, a mass spectroscopy apparatus or data analysis software and
XX
XX (b) treating the fractionated biological samples and the reference
XX
XX samples with the gas chromatography apparatus to generate chromatographic
XX
XX data corresponding to the fractionated biological samples and the
XX
XX reference samples, (c) treating the fractionated biological samples and
XX
XX the reference samples with the mass spectroscopy apparatus to generate
XX
XX spectroscopic data corresponding to the fractionated biological samples
XX
XX and the reference samples and (d) processing the chromatographic and the
XX
XX spectroscopic data with the data analysis software. The nucleic acid and
XX
XX the vector are useful for altering the metabolism of a plant and for
XX
XX stunting a plant. The nucleic acids are useful in agriculture to create
XX
XX dwarf varieties of any plant species. The present sequence is a
XX
XX plant cDNA contig or singleton (related to dwarfism/stunting) identified
XX
XX by searching a nucleic acid database with plant EST (expressed
XX
XX sequence tag) and a BLAST (basic local alignment tool) stringency
XX
XX of e-20.
SQ
Sequence 663 BP; 150 A; 204 C; 150 G; 159 T; 0 other;
Query Match 3.0%; Score 44.8; DB 24; Length 663;
Best Local Similarity 50.8%; Pred. No. 0.0023;
Matches 134; Conservative 0; Mismatches 127; Indels 3; Gaps 1;
QY 316 CACCTGTTCACTATGAGGCTGAGCCAGCTTGCAGCCCTACCTGCTGATGCTGAC 375
DB 250 CTCCTCTCAAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 309

```

QY 376 TTGATGTGTCCTGCCCCGAGAAGGCTGGAGGTGCCCCCATTC--TCTGGGTG 432
    ||||| || ||| ||||| ||||| ||||| ||||| |||||
Db 310 ACCGATGTCTTCCTTCGAGACTCCAAATGACTCACCATCCGCTCCAAGCTCACATG 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 GACCGTATGGCGCTCATCTATGTCGGGCACACTGACGACACAAGACTCTGTGATGCA 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 GACCACATGGCGACATCTATGCGCAGGGTTCCCAAGACATGAAATGCGTGGGATGCA 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 493 TTACTGAGGCGCTTGGAGCTCTCTGTGATCAGAGTACATCCCCGAAAGATCTTCTTC 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 TACCTGAGGCGCATACGACAGCTTCCTTGCTCAAGCCACTCCGATCCGTCTAT 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 ATTCTCTGGGCGCATGATGAGAG 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 CTCTCTCTGTCGCCCGATGAGAG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: July 1, 2003, 05:58:54
 Job time : 406 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 04:22:44 ; Search time 2340 seconds

(without alignments)
10444.013 Million cell updates/sec

Title: US-10-014-896-1

Perfect score: 1509
1 atggctcagcggtgctgttg.....ctcacctgcacaaactgtga 1509Sequence: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Scoring table: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.*
1: em_estlba:*
2: em_estlhum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlrv:*
6: em_estlsc:*
7: em_estlro:*
8: em_estlhc:*
9: gb_estl1:*
10: gb_estl2:*
11: gb_estl3:*
12: gb_estl4:*
13: gb_estl5:*
14: gb_estl6:*
15: em_estlun:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hiv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	737.8	48.9	741	12	BG697810 602661011
2	726.4	48.1	773	12	BG740058 602631131
3	724.2	48.0	805	12	BG741998 602633443
4	695.8	46.1	769	12	BG742908 602632491
5	693.2	45.9	970	12	BG674119 602619926
6	690.2	45.7	771	12	BG675035 602621268

7	532.6	35.3	1023	12	BG743823	BG743823 602632667
8	529.4	35.1	847	13	B1143597	B1143597 602907430
9	528.4	35.0	575	13	B1832659	B1832659 603082042
10	518.4	34.4	853	13	B1218153	B1218153 602934270
11	465.2	30.8	706	13	B1101355	B1101355 602887103
12	459.2	30.4	851	12	BF234135	BF234135 602024884
13	424.2	28.1	748	12	BF234647	BF234647 602028464
14	421.6	27.9	666	10	BB613230	BB613230 BB613230
15	410.2	27.2	580	12	BG608571	BG608571 307199 MA
16	406.2	26.9	580	10	BE032236	BE032236 131103 MA
17	404.2	26.8	614	12	BF525111	BF525111 UI-R-AC07
18	401.6	26.6	544	12	BG608554	BG608554 307177 MA
19	399.4	26.5	637	10	BB619403	BB619403 BB619403
20	396	26.2	612	13	B1219686	B1219686 602935391
21	394.4	26.1	565	10	BE032261	BE032261 131134 MA
22	391.4	25.9	564	13	BM088240	BM088240 501660 MA
23	385.6	25.6	561	12	BF079682	BF079682 230350 MA
24	385.6	25.6	584	13	B1339966	B1339966 365091 MA
25	377.6	25.0	576	13	B1339967	B1339967 365091 MA
26	350	23.2	849	12	BF533743	BF533743 602073955
27	340	22.5	717	13	B1147849	B1147849 602912738
28	313	20.7	822	13	B1148134	B1148134 602912305
29	308.6	20.5	567	12	BG895222	BG895222 358510 MA
30	277.2	18.4	1267	12	BG674056	BG674056 602620050
31	252.2	16.7	338	10	AM663138	AM663138 bh75b07.Y
32	250.4	16.6	711	14	BE063860	BE063860 MT-P-CP1-
33	247	16.4	430	10	BB846588	BB846588 BB846588
34	245.4	16.3	385	10	BB872723	BB872723 BB872723
35	239.2	15.9	742	13	B1247973	B1247973 602958833
36	224.6	14.9	617	13	BM440197	BM440197 P911n.PKD
37	223.2	14.8	737	9	AJ456300	AJ456300 AJ456300
38	220	14.6	368	10	BB871294	BB871294 BB871294
39	195	12.9	378	10	BB872712	BB872712 BB872712
40	194.8	12.9	350	10	BB869217	BB869217 BB869217
41	192.6	12.8	321	9	A1317512	A1317512 u123f05.Y
42	192.6	12.8	416	9	A1317514	A1317514 u123f07.Y
43	188.8	12.5	533	14	B0074284	B0074284 f227a06.Y
44	186.6	12.4	490	17	BH044325	BH044325 RPICT-24-12
45	177.4	11.8	544	13	B1672970	B1672970 ft40b08.Y

ALIGNMENTS

RESULT 1
LOCUS BG697810 741 bp mRNA
DEFINITION 602661011F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4004207.5,
mRNA sequence.
ACCESSION BG697810 GI:13964441
VERSION BG697810.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcgaps-remail.nih.gov

Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.lnl.gov
Plate: LHAM10701 Row: e Column: 03
High quality sequence start: 4
High quality sequence stop: 732.
Location/Qualifiers

FEATURES

QY 481 TCGTATGATGATTTACTGAGAGGCTTGAGAGTCTCTGATGATGAGAAATACATCCCGA 540
|||||
Db 526 TCTGTATGATGATTTACTGAGAGGCTTGAGAGTCTCTCTATGATGAGAAATACATCCCGA 585
QY 541 AGATCTTCTGATTTCTCTGAGGCGCATGATGAGAGATCATGAGGAGGAGGCTCAGAG 600
|||||
Db 586 AGATCTTCTGATTTCTCTGAGGCGCATGATGAGAGATCATGAGGAGGAGGCTCAGAG 645
QY 601 ATTCAGAGGCTGCTAGAGTCAAGAGGCGCTCAGAGTACCTTATGATGAGAGGAGGCG 660
|||||
Db 646 ATTCAGAGGCTGCTAGAGTCAAGAGGCGCTCAGAGTACCTTATGATGAGAGGAGGCG 705
QY 661 TTCTATCTGATGATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
|||||
Db 706 TTCTATCTGATGATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 765
QY 721 GAGAGAGG 728
|||||
Db 766 GAGAGAGG 773

RESULT 3
BG741998 805 bp mRNA linear EST 15-MAY-2001
LOCUS 60263443F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778507 5',
DEFINITION mRNA sequence.
ACCESSION BG741998
VERSION BG741998.1 GI:14052651
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 805)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9abps-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10634 row: f column: 12
High quality sequence stop: 771.
Location/Qualifiers
1. 805

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4778507"
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 184 a 211 c 225 g 185 t
ORIGIN

Query Match 48.0%; Score 724.2; DB 12; Length 805;
Best Local Similarity 98.2%; Pred. No. 1.8e-187;
Matches 743; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 1 ATGGCTCAGAGGCTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
|||||
Db 46 ATGGCTCAGAGGCTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 105
|||||
QY 61 ACCGCTCCAGATGATGAGGCGCGAGAGGCGGAGCATCAAGAGGCGTCCGAATCCCT 120
|||||

Db 106 ACCGCTCCAGATGATGAGGCGCGAGAGGCGGAGATATCAAGAGGCGTCCGAATCCCT 165
QY 121 TCTCAGTTACAGAAAGAGAAAGCGCTGCGATGAAAGAGGCGTGAAGAGTCCATCCAG 180
|||||
Db 166 TCTCAGTTACAGAAAGAGAAAGCGCTGCGATGAAAGAGGCGTGAAGAGTCCATCCAG 225
QY 181 ATTCCAAAGTATGATTTTACCTGAGAGTCCAAATATATACAGCCCTGCTGAGTTCGGA 240
|||||
Db 226 ATTCACAGATGATTTTACCTGAGAGTCCAAATATATACAGCCCTGCTGAGTTCGGA 285
QY 241 AATACATTCATTAAGCTTTTCTGAGAGTCCAAATATATACAGCCCTGCTGAGTTCGGA 300
|||||
Db 286 AATACATTCATTAAGCTTTTCTGAGAGTCCAAATATATACAGCCCTGCTGAGTTCGGA 345
QY 301 GTGAAAGATATAGCACCTGCTTCAATCCAAAGGCTGAGACCCAGCTTACAGCCCTAC 360
|||||
Db 346 GTGAAAGATATAGCACCTGCTTCAATCCAAAGGCTGAGACCCAGCTTACAGCCCTAC 405
QY 361 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
|||||
Db 406 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465
QY 421 TTCTCTGAGTGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
|||||
Db 466 TTCTCTGAGTGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 525
QY 481 TCTGTATGATGATTTACTGAGAGGCTTGAGAGTCTCTGATGATGAGAAATACATCCCGA 540
|||||
Db 526 TCTGTATGATGATTTACTGAGAGGCTTGAGAGTCTCTGATGATGAGAAATACATCCCGA 585
QY 541 AGATCTTCTGATTTCTCTGAGGCGCATGATGAGAGATCATGAGGAGGAGGCTCAGAG 600
|||||
Db 586 AGATCTTCTGATTTCTCTGAGGCGCATGATGAGAGATCATGAGGAGGAGGCTCAGAG 645
QY 601 ATTCAGAGGCTGCTAGAGTCAAGAGGCGCTCAGAGTACCTTATGATGAGAGGAGGCG 660
|||||
Db 646 ATTCAGAGGCTGCTAGAGTCAAGAGGCGCTCAGAGTACCTTATGATGAGAGGAGGCG 705
QY 661 TTCTATCTGATGATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
|||||
Db 706 TTCTATCTGATGATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 765
QY 721 GAGAGAGG-TACATGAACTCATGCTGAGAGTAAACA 757
|||||
Db 766 GAGAGAGG-TACATGAACTCATGCTGAGAGTAAACA 801

RESULT 4
BG742908 769 bp mRNA linear EST 15-MAY-2001
LOCUS 6026342491F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4777726 5',
DEFINITION mRNA sequence.
ACCESSION BG742908
VERSION BG742908.1 GI:14053561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 769)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9abps-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10632 row: e column: 23

High quality sequence stop: 767.

FEATURES

Location/Qualifiers

1..769

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4777726"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT

211 a 221 c 169 g 168 t

ORIGIN

Query Match 46.1%; Score 695.8; DB 12; Length 769;
Best Local Similarity 99.7%; Pred. No. 1.1e-179;
Matches 697; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 811 GCAGCTGCTGACCCGATTTGAGACAGACACCAATGCTATCATATTTGGAGCGGAGCA 870
DB 1 GCAGCTGCTGACCCGATTTGAGACAGACACCAATGCTATCATATTTGGAGCGGAGCA 60
DB 61 GTGTGACTGTATTTGAGACAGACCAATGAGTTCCTTCCCTGCAATATATATCTG 120
QY 931 AGCAACCCATGCTATTTGAGACCACTTATAGAGGTTTATGAGAGAAATCCCTTACC 990
DB 121 AGCAACCCATGCTATTTGAGACCACTTATAGAGGTTTATGAGAGAAATCCCTTACC 180
QY 991 AATGCAATATATGAGACAGACAGGACACTACCATATTTCAAGCAGGGGTCAAGTCAAT 1050
DB 181 AATGCAATATATGAGACAGACAGGACACTACCATATTTCAAGCAGGGGTCAAGTCAAT 240
QY 1051 GTATATCCCTGACCTGAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 1110
DB 241 GTATATCCCTGACCTGAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 300
QY 1111 CAAGAGGCTCTAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 1170
DB 301 CAAGAGGCTCTAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 360
QY 1171 TTGAGTCCCTTTGAGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1230
DB 361 TTGAGTCCCTTTGAGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 420
QY 1231 CTGCTCCGCGAGACCTGACAGTCCGTCCTCCGGAAGTCAATTTACTGCCCACTTACT 1290
DB 421 CTGCTCCGCGAGACCTGACAGTCCGTCCTCCGGAAGTCAATTTACTGCCCACTTACT 480
QY 1291 TCTATGCGACAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 1350
DB 481 TCTATGCGACAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGAC 540
QY 1351 TACCCATCTATACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 1410
DB 541 TACCCATCTATACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 600
QY 1411 TTAGTTCAGAGCTATGAGACCAAGTAAATTCATTTGAGTTTGAATTCAGATCTGAC 1470
DB 601 TTAGTTCAGAGCTATGAGACCAAGTAAATTCATTTGAGTTTGAATTCAGATCTGAC 660
QY 1471 ACAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 1509
DB 661 ACAGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 699

RESULT 5
Bg674119 970 bp mRNA linear EST 01-MAY-2001
LOCUS Bg674119 602619926p1 NCI_CGAP_Skn3 Homo sapiens cDNA IMAGE:4745407 5',
DEFINITION mRNA sequence.

ACCESSION Bg674119
VERSION Bg674119.1 GI:13905515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 970)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgep@b-rs-mail.nih.gov

Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN). DNA
Sequencing by: Invitae Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: LLM10592 row: c column: 08
High quality sequence stop: 695.

FEATURES

Location/Qualifiers

1..970

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4745407"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT

269 a 254 c 228 g 219 t

ORIGIN

Query Match 45.9%; Score 693.2; DB 12; Length 970;
Best Local Similarity 91.2%; Pred. No. 6.6e-179;
Matches 872; Conservative 0; Mismatches 63; Indels 21; Gaps 12;

QY 486 GATGGCATTACTGACAGGCTTGAGAGCTCTGATGACAGGACAGGACAGGACAGGACAGG 545
DB 1 GATGGCATTACTGACAGGCTTGAGAGCTCTGATGACAGGACAGGACAGGACAGGACAGG 60
QY 546 TTTCTTCTTCTCTGAGGACATGATGAGAGATCATGAGGACAGGACAGGACAGGACAGG 605
DB 61 TTTCTTCTTCTCTGAGGACATGATGAGAGATCATGAGGACAGGACAGGACAGGACAGG 120
QY 606 AGCCCTGTACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 665
DB 121 AGCCCTGTACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 180
QY 666 CTGGATGATTTTCAATCTTCACTTCAAGAAAGCCATGCTTGTATGACAGTCAAGAA 725
DB 181 CTGGATGATTTTCAATCTTCACTTCAAGAAAGCCATGCTTGTATGACAGTCAAGAA 240
QY 726 GGGTTCATGAACTTCACTGCTCAAGTAAACATGACTTCAAGGACAGGACAGGACAGG 785
DB 241 GGGTTCATGAACTTCACTGCTCAAGTAAACATGACTTCAAGGACAGGACAGGACAGG 300
QY 786 AAAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 845
DB 301 AAAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 360
QY 846 GCTATCATATTTGAGAGCGGAGAGGAGTGTGATGATGACAGCACTGGCAATAGATG 905
DB 361 GCTATCATATTTGAGAGCGGAGAGGAGTGTGATGATGACAGCACTGGCAATAGATG 420
QY 906 TCCCTTCCCTGTCAA-TATATCTGAGCAACCATGGCTATTGAACTTATTAAGCA 964
DB 421 TCCCTTCCCTGTCAA-TATATCTGAGCAACCATGGCTATTGAACTTATTAAGCA 480
QY 965 GGTATTAGAGAGAAATCCCTTAACCAATCAATATCAAGACCAACGAGGACAGCACTGACCA 1024

Db	481	GGTTTATGGAGAGAAATCCCTTAACCAATGCAATTAATCAGGACACACCGGCACTGACCA	540
QY	1025	TATTTCAAAGCAGGGGGGTCAAGTTCAATGTCTATC--CCGCCAGTGGCCCGCAGGCAAGTCAA	1082
Db	541	TATTTCAAAGCAGGGGGGTCAAGTTCAATGTCTATCACCACCCAGTTGGCCGACAGGCAAGTCAA	600
QY	1083	CTTCCGGATTCACCCCTGGACAGACAGTCCCAAGAGTTC--TAGAACTACGAAAGAACATTG	1141
Db	601	CTTCCGGATTCACCCCTGGACAGACAGTCCCAAGAGTTCCTTAGAACTACGAAAGAACACTG	660
QY	1142	TGGCTGATTAACAGATGCC-AGTTCACATGTGTAGAGT--CCCTTGGACCCCTCCCGGTCA	1198
Db	661	TGGCTGATTAACAGAGTCCAAAGTTCACATGTGTAGAGTGGCCTTTGAAACACATCCCGGTCA	720
QY	1199	GCCCTTCT---GATGACAAAGGCGTTGGG-----CTACACAGCTGCTCCGGCAGACCGTAC	1249
Db	721	GACCATATCTGGAATGACAAAGGCGCTTGGGGCTTAACAGACTGGCTCCGGCCAGAGCTGTACCA	780
QY	1250	AGTCCGCTCTTCCCGGAAGTCAATA-TTACTGCCCCAGTTACTTCTATTATGGCAACACAG-A	1307
Db	781	GTCGGGCTTCCCGGAAGTCAATATTACTGCGCCCAATACTAATTATTGCAACACAGAA	840
QY	1308	CAGCCGATTTCTTTCACAAACTCA-CCACTGGCATCTACAG-GTTCTACCCCATCTACATA	1365
Db	841	CAGACGATTTCACTTAAAAACTCACAGCGATGGCGATGTACAGAGTTCGTATCCCATCTACATA	900
QY	1366	CAGCCTGAAGCTTCAAAAGCATTCAGTGAAGTCAACGAGAAATCTCAAGTCCAGG	1421
Db	901	CAGCC-GGAGACTTTGAAAGGCATTCATGAGATCAACGACAACTTATGGCCAGGC	955

FEATURES	source
LOCUS	Bg675035
DEFINITION	Bg675035 771 bp mRNA linear EST 01-MAY-2001
VERSION	6026212268F1
KEYWORDS	NCI_CGAP_Skn3 Homo sapiens CDNA clone IMAGE:4746689 5', 3'
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 771)
TITLE	NH-MGC http://mgc.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgep@rs-remail.nih.gov Tissue Procurement: James Cleaver, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLM10595 row: h column: 18 High quality sequence stop: 750.
FEATURES	Location/Qualifiers
source	1..771
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:4746689"
	/clone_11b="NCI_CGAP_Skn3"
	/lab_host="DH10B (TI phage-resistant)"
	/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NCI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo df. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT	170 a 203 c 220 g 176 t
ORIGIN	

[illegible]

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bms-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNL0632 row: f column: 18
High quality sequence stop: 749.
Location/Qualifiers
1. 1023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="477745"
/clone_id="NCL_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT 281 a 279 c 223 g 240 t
ORIGIN
Query Match 35.3%; Score 532.6; DB 12; Length 1023;
Best Local Similarity 99.1%; Pred. No. 8.4e-135;
Matches 546; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 960 AAGCAGGTTTATGAGAGAAATCCCTTACCAATGCAATATACAGACCCAGGCACT 1019
DB 14 AATGAGGTTTATGAGAGAAATCCCTTACCAATGCAATATACAGACCCAGGCACT 73
QY 1020 CACCATATTCAGAGAGGCTCAAGTTCATATGTCATCCGCCAGTGGCCAGGCAAGT 1079
DB 74 CACCATATTCAGAGAGGCTCAAGTTCATATGTCATCCGCCAGTGGCCAGGCAAGT 133
QY 1080 CACTTCCGAGTTCACCTGGAGACAGAGTCCAGAGAGTCCAGAGAGTCCAGAGAGT 1139
DB 134 CACTTCCGAGTTCACCTGGAGACAGAGTCCAGAGAGTCCAGAGAGTCCAGAGAGT 193
QY 1140 TGTGCGATTAACAGAGTCCAGTTCATATGAGTGGCTTGGACCCCTCCCGTCA 1199
DB 194 TGTGCGATTAACAGAGTCCAGTTCATATGAGTGGCTTGGACCCCTCCCGTCA 253
QY 1200 CCTTCTGATGACAGAGGCTTGGCTACAGAGTCCCGCAGAGAGTCCCGTCTT 1259
DB 254 CCTTCTGATGACAGAGGCTTGGCTACAGAGTCCCGCAGAGAGTCCCGTCTT 313
QY 1260 CCGGAGTTCATATTAATG-CGCCAGTTCATATGAGTGGCAACAGAGAGGCAATCT 1318
DB 314 CCGGAGTTCATATTAATG-CGCCAGTTCATATGAGTGGCAACAGAGAGGCAATCT 373
QY 1319 TTACAAACCTCAGCAGTTCAGAGTTCATATGAGTGGCAACAGAGAGGCAATCT 1378
DB 374 TTACAAACCTCAGCAGTTCAGAGTTCATATGAGTGGCAACAGAGAGGCAATCT 433
QY 1379 TCAACGAGTTCAGAGTTCAGAGTTCATATGAGTGGCAACAGAGAGGCAATCT 1438
DB 434 TCAACGAGTTCAGAGTTCAGAGTTCATATGAGTGGCAACAGAGAGGCAATCT 493
QY 1439 AATTCATCTTGAAGTTCAGAGTTCAGAGTTCATATGAGTGGCAACAGAGAGG 1498
DB 494 AATTCATCTTGAAGTTCAGAGTTCAGAGTTCATATGAGTGGCAACAGAGAGG 553
QY 1499 ACAACCTGTGA 1509
DB 554 ACAACCTGTGA 564
RESULT 8
B1143597
LOCUS B1143597 847 bp mRNA linear EST 05-JUL-2001

DEFINITION 602907430F1 NCL_CGAP_K1d14 Mus musculus cDNA clone IMAGE:5064342
5', mRNA sequence.
ACCESSION B1143597
VERSION B1143597.1 GI:14603598
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS 1 (bases 1 to 847)
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bms-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNL01173 row: 1 column: 07
High quality sequence stop: 806.
Location/Qualifiers
1. 847
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="5064342"
/clone_id="NCL_CGAP_K1d14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library. |"
BASE COUNT 193 a 239 c 222 g 193 t
ORIGIN
Query Match 35.1%; Score 529.4; DB 13; Length 847;
Best Local Similarity 81.5%; Pred. No. 5.6e-134;
Matches 674; Conservative 0; Mismatches 146; Indels 7; Gaps 5;
QY 1 AATGCTCAGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 59
DB 3 AATGCTCAGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62
QY 60 TACGCTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 119
DB 63 TACGCTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 122
QY 120 TTTCAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 179
DB 123 TTTCAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 182
QY 180 GATTCACAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 239
DB 193 GATTCACAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 242
QY 240 AATATCATTAATTAAGTTCCTTACAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 299
DB 243 AATATCATTAATTAAGTTCCTTACAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 302
QY 300 CGTGAAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 359
DB 303 CGTGAAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 362
QY 360 CCTGCTGATGCTCATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
DB 363 CATGCTGATGCTCATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
QY 420 AATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479

Db	427	TTCTCAGGCGCTGGAAAGCATATGGCTTCATCTATATGCGCGGGGTGGCTGGACAACAAAAAAC	486
Qy	481	TCCTCATATGCATTAATCGACAGCGCTTGAGACTCTCTGTGATCAGAGAGTACATCCCCGA	540
Db	487	TCCTGATGGGAGATCCTCGATAGCTTTGGAGTCTCTGTGATCAGAAACTACAGCCCAA	546
Qy	541	AGATCTTTCTCA-TTTCCTGGGCGCATGATGAGAGTCAATCAGG---ACAGGGGCTC	595
Db	547	AGACTCTTCTCACTTCTCTTGGGCGCATGATGAGAGAGTGTCCGGGAAAAAGGGGGCTC	606
Qy	596	AGAGATCTAGCCCTCTCAAGTCAAGCAAGGGGCGTCCAGCTAGCTTATTTGGACGAG	655
Db	607	AGAAAGATCTAGCACTTTACAGGCAAGGGGGTGTCCAGCTAGCTTCTTGTGATGAAG	666
Qy	656	GGGCTTCATCTTGATGATTTCACTTCTCACTTCAACA	695
Db	667	GGAGCTTATCTTGAAGGCTTCACTTCAAAAGCTTCGAGAA	706

RESULT	12
LOCUS	BF234135
DEFINITION	BF234135 851 bp mRNA linear EST 14-NOV-2000 60202888.f1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4160300 5'
ACCESSION	BF234135
VERSION	BF234135
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 851)	NIH-MGC	http://mgc.nhl.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.			

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M9439 row: 9 column: 21
High quality sequence start: 2
High quality sequence stop: 722.
Location/Qualifiers

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image:4160300"
/clone_id="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT      194 a      237 c      229 g      191 t
ORIGIN

```

```

Query Match 30.4%; Score 459.2; DB 12; Length 851;
Best Local Similarity 78.3%; Pred No. 1,1e-114;
Matches 617; Conservative 0; Mismatches 156; Indels 13; Gaps 5

QY      1  ATGGCTCAGCGGTGGCTTGGCTGCTGGCCCTGATGGCTATGCTCTCTACTATTTCCCT 60
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       37  ATGGCTGAGCTACTTGTAGCTTTGGCCCGCGCTGGCAGCTGTGCTCTCTTTTTCGT 96
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      61  ACCGCTCCAGTTCATGGGCCCCGAGAGCGGGGAGCATCAAGGGCGTCGGATTCCT 120
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	97	ACGGCTCTCCGGATCCACATGGGCCCTTAGAAGCGGGGAAAAATCGG66GGGGCTCCGGATCCCT	156
OY	121	TTCTAGTTTCACCAAGAGAGACGGGTGGCCATGTGAAAAGGGCGTGAAGAGTCCATCCAG	180
Db	157	TTCCAGTTCACCGAGGAGAGCGTGTGCTCTATAAAGGGCGCTGAAGAAGGTCCATCCAG	216
OY	181	ATTCCACAGTGAATCTTTTAAAGTCTTGAGAAAGTCCATATCTACAGCCCTTGCTGAGTTGGA	240
Db	217	ATTCCACAGTGTCTTTTCAGCCAGAGAGAAATCCAAACACACAGCCCTTGCTGAGTTTGA	276
OY	241	AAATACATTCATAAAGTCTTTCCTACAGTGGTCAACACACACTTATCCAGCTGAATC	300
Db	277	GAAATATATCCCGAAAGCCCTTCCCTACATAGTTTCCACACACACCTTGTCACATGAATTC	336
OY	301	GTGGAAGAGTATPACCCACACTGTTTACATATCCATGCAAGGCTGGAGCCACGTTGCTCAGCCCTAC	360
Db	337	GTGGCAAGTATAGCACACACTGTTTACACATCCATCCAAAGGCTCAGACCCAGTTTGACGCCCTAC	396
OY	361	CTGCTGATGGCTCACTTGTATGTGTGGTGGCTGCCCTGAAGAAGCTGGGAAGTGTCCCCCA	420
Db	397	ATGCTGATGAGGTCCACATTTGATGTGGTTCCTCCCGGAAGAGATGGGAGTGGCCCCG	456
OY	421	TTTCTGTGGGTTGAGAGCGTATGGCGTCATGTCGTGGGACACACTGAGCAGCAAGAAC	480
Db	457	TTCTCAGGCCCTGGAAGCGCAATGGCTTATCTATAGCCGGGGGCTGTGACACAAAAAC	516
OY	481	TTCTGATGCAATTACTGCAGAGCCTTGGAGCTCTGCTGATCAGGAGTACATCCCCGA	540
Db	517	TCTGTGAAGGGGATCTCTCATG-CCTTGGAGCTCCTGTTGATCAGAACTACAGCCCAAA	576
OY	541	AGATTTTCTTCATTTCTGTGGGCCATGATGAGAGTATCAGG-ACAGGGGGTCT	596
Db	576	AGATCTTCTTCATTTGCTTTGGGCCATGATGAGAGGTGTCCGCGGGAAGAAAAAGGGGCTC	636
OY	596	AGAGATCTACAGCCCTGCTACAGTCAAGGGGGGT-CCAGCTAGCTTCATTTGTGACGAG	654
Db	636	AGAAATCTTCAGCACTCTTACAGGCAAGGGGTGTCCAGCTAGCCCTTCTTGTGATGAA	696
OY	655	GGGGGCTTCATCTTGGATGATTTCACTTCACTTCAAGAACCCCATGCGCTTGATTTGA	714
Db	696	GGGAGCTTCATCTTGGAGGCTTCAATTCACAAACCTCGAGAAAG-CCGTTGCCATGATTCAG	754
OY	715	GTCTCAGAGAAAGGTTTCATGAACCTCATGTGCAAGTAAACATGACTTCAGGCCACTCT	774
Db	755	CACGTACGAAAGGTTCTTGACCTGCTGCGAG-----TAAATGACTTCAGGCCACTCT	809
OY	775	TCAGCTCC 782	
Db	810	TCAGTCC 817	

RESULT	13
LOCUS	BF234647
DEFINITION	BF234647 748 bp mRNA EST 14-NOV-2000 602028864F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4163942 5'
ACCESSION	BF234647
VERSION	BF234647.1 GI:11146356
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 748)
TITLE	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc.

Tue Jul 1 11:49:16 2003

us-10-014-896-1.rst

Page 12

Db 419 TTGGAGTCTCCGAAAGGGCTTAATTAACCTATGCTCCAACTAAACACGATCTCAGGCC 478

Qy 770 ACTCTTCAGTCTCTCCAAAGGAGACAGCATTTGGCATCTTGCAGCTGCTGCACGCCAT 829

Db 479 ACTCTTCAGTCTCTCCAAAGGAACTAGCATTTGGTATCTTACGACGCGCTGTCAACGAC 538

Qy 830 TGGAGCAGACACCAATGCTATCATATTGGAAGGGGACA 870

Db 539 TGGACAGGACCAATGCCAACAATCATTTGGAGATGGGCCA 579

Search completed: July 1, 2003, 07:46:16
Job time : 2359 secs

query match 3.38; Score 49.2; DB 1; Length 1224

Best Local Similarity 51.9%; Pred. No. 1.2e-05;
Matches 136; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

```

QY 318 CCTGTACTATCCAGGCTCGACGCCAGCTTGCACCCCTACCTGTGATGGCTACTT 377
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CGTTTACCTGGCCAGCACCACCTTACCTCTCTCATCTGTCACATCCACAC 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 378 TGAATGGTGGCTGCCCCCTGAAGAGGCTGGAGTCCGCCATCTCTGGGT---GGA 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GGATGGTGGCTGCTCTTCAAGGAACTGGAGTCACGACCCCTTGAAGGCTTCAAGGA 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 435 GCGTATGGGCTCATCTATGTGGGGCCACCTGAGCAAGAACCTCTGTATGGCATT 494
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 TTCTGAGGGCTACTATCTATGCGAGGGTCCAGGACATGATGCGTCAGATCCAGTA 362
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 495 ACTGAGGCTTGGAGCTCTGCTGATCAGGAATACATCCCGAAGATCTTCTTCAT 554
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 CCTGAGAGCTGTGAGGAGGCTGAAGTGGAGGGCCACCGGTTCGCCAAGAACATCCACAT 422
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 555 TTCTCTGGCCATGATGAGAG 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 GACCTTGTGCTGATGAGAG 444
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2

US-08-555-860-3
Sequence 3, Application US/08555860
Patent No. 5585474

GENERAL INFORMATION:

APPLICANT: IMAKI, Kanso
APPLICANT: OHRA, Tsunetaka
APPLICANT: KURIHARA, Masahiro
TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/555,860
FILING DATE: 13-NOV-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/127,278
FILING DATE: 27-SEP-1993
APPLICATION NUMBER: JP 281136/1992
FILING DATE: 28-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: NEIMARK, Sheridan
REGISTRATION NUMBER: 20,520
REFERENCE/DOCKET NUMBER: IMAKI-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 1224 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

LOCATION: 1..1224
US-08-555-860-3

Query Match 3.3%; Score 49.2; DB 1; Length 1224;
Best Local Similarity 51.9%; Pred. No. 1.2e-05;
Matches 136; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

```

QY 318 CCGTTTCACTATCCAGGCTCGACGCCAGCTTGCACCCCTACCTGTGATGGCTACTT 377
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CGTTTACCTGGCCAGCACCACCTTACCTCTCTCATCTGTCACATCCACAC 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 378 TGAATGGTGGCTGCCCCCTGAAGAGGCTGGAGTCCGCCATCTCTGGGT---GGA 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GGATGGTGGCTGCTCTTCAAGGAACTGGAGTCACGACCCCTTGAAGGCTTCAAGGA 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 435 GCGTATGGGCTCATCTATGTGGGGCCACCTGAGCAAGAACCTCTGTATGGCATT 494
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 TTCTGAGGGCTACTATCTATGCGAGGGTCCAGGACATGATGCGTCAGATCCAGTA 362
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 495 ACTGAGGCTTGGAGCTCTGCTGATCAGGAATACATCCCGAAGATCTTCTTCAT 554
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 CCTGAGAGCTGTGAGGAGGCTGAAGTGGAGGGCCACCGGTTCGCCAAGAACATCCACAT 422
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 555 TTCTCTGGCCATGATGAGAG 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 GACCTTGTGCTGATGAGAG 444
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 3

US-09-814-951A-1
Sequence 1, Application US/09814951A
Patent No. 6387661

GENERAL INFORMATION:

APPLICANT: SHAO, Wei et al
TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
FILE REFERENCE: CLO01179
CURRENT APPLICATION NUMBER: US/09/814,951A
CURRENT FILING DATE: 2001-03-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1445
TYPE: DNA
ORGANISM: Homo sapiens

US-09-814-951A-1

Query Match 3.0%; Score 46; DB 4; Length 1445;
Best Local Similarity 51.6%; Pred. No. 0.00015;

Matches 131; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

```

QY 326 CTATCCAGGCTCGACGCCAGCTTGCACCCCTACCTGATGGCTGATGGTATGG 385
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 CTGACTATGGCAGCACCACCTTACCTCTCTCATCTGTCACATCCACACGATGG 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 TGCCTGCCCTGGAAGAGGCTGGAGGTGCCCCCATCTCTGGGT---GGAAGTGAAT 442
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 TGCCTGTCTTCAAGGAACTGGAGTCACGACCCCTTGAAGGCTTCAAGGATTCGAG 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 443 GCGTCATCTATGTCGGGGCCACCTGAGCAGCAAGAACCTGTGATGGCAATTAAGCAG 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 GCTACATCTATGCGAGGGGTGCCAGGACATGAAGTCCGACATCCACTACTGGAAG 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 503 CCTGAGAGCTCTGCTGATCAGGAATACATCCCGAAGATCTTCTTCTATCTCTG 552
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 382 CTGTGAGGAGCTTAAGTGGAGGGCCACCGGTTCGCCAAGAACATCCATGACCTTGG 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 563 GCCATGATGAGAG 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 TGCCTGATGAGAG 455
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4

US-08-204-740-5
Sequence 5, Application US/08204740
Patent No. 5753432
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrei
APPLICANT: Kazarov, Alexander
APPLICANT: Mazo, Ilya
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods for Identifying Genetic
TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
TITLE OF INVENTION: Growth in Cancer Cells
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allgretti & Witcoff, Ltd.
STREET: 10 S. Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/204,740
FILING DATE: 04-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5753432nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-204-740-5
Query Match 2.5%; Score 37.4; DB 1; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.032; 71; Indels 2; Gaps 1;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;
QY 315 CCACCTGTTCACTATCCAGGCTCGGACCCCAAGCTTGCACTGCTGAGTGCCTCA 374
DB 108 CCATCTGCTGGGCAAGTAGGACGACGCCCAAGAAACCGTGTGCTTACGGGCA 167
QY 375 CTTTATGATGCTGCT--GCCCTGAGAGAGGCTGGGAGGTCGCCCATTCCTGCGGTG 432
DB 168 CTTGACGCTGCGCCTGCGCCCTGGAGAGAGGCTGGACAGGACCCCTTACCTTGGTG 227
QY 433 GAGCGTGAAGGCGCTGATGCTGCGGCGACACTGACGACAAAG 477
DB 228 GAGCGGAGAGCAACTGTATGGAGAGGCTCCACGACGATTAAG 272
RESULT 5
US-09-081-167A-5
Sequence 5, Application US/09081167A
Patent No. 6083745
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrei
APPLICANT: Kazarov, Alexander
APPLICANT: Mazo, Ilya
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods for Identifying Genetic
TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

TITLE OF INVENTION: Growth in Cancer Cells
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 S. Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,167A
FILING DATE: 18-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6083745nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-KK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-081-167A-5
Query Match 2.5%; Score 37.4; DB 3; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.032; 71; Indels 2; Gaps 1;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;
QY 315 CCACCTGTTCACTATCCAGGCTCGGACCCCAAGCTTGCACTGCTGAGTGCCTCA 374
DB 108 CCATCTGCTGGGCAAGTAGGACGACGCCCAAGAAACCGTGTGCTTACGGGCA 167
QY 375 CTTTATGATGCTGCT--GCCCTGAGAGAGGCTGGGAGTGCCTTCTGCGGTG 432
DB 168 CTTGACGCTGCGCCTGCGCCCTGGAGAGAGGCTGGACAGGACCCCTTACCTTGGTG 227
QY 433 GAGCGTGAAGGCGCTGATGCTGCGGCGACACTGACGACAAAG 477
DB 228 GAGCGGAGAGCAACTGTATGGAGAGGCTCCACGACGATTAAG 272
RESULT 6
US-09-081-395-5
Sequence 5, Application US/09081395
Patent No. 6083746
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrei
APPLICANT: Kazarov, Alexander
APPLICANT: Mazo, Ilya
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods for Identifying Genetic
TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
TITLE OF INVENTION: Growth in Cancer Cells
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 S. Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081.395
FILING DATE: 18-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6083746nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-KK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-081-395-5

Query Match
Best Local Similarity 2.5%; Score 37.4; DB 3; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.032;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

QY 315 CCACCTGTTCACTATCCAGAGCTGGACCCAGCTTGACCCCTACTGCTGATGCTCA 374
DB 108 CCATCTGCTGGGCAAGCTAGGACGACCCCAAGAAACCGTGCTGATTTACGGGCA 167
QY 375 CTTTGATGTGGTGCT--GCCCTGAAGAGGCTGGAGGTCGCCCATTTCTGTGGTTG 432
DB 168 CCTGGAGCTGACAGCTGCGCCCTGGAGAGCGGGTGGAGACGACGACCCCTTACCTTGGTG 227
QY 433 GAGCGTATGGCGCTCATCTATGTCGGGGCACACTGACGACAAAG 477
DB 228 GAGCGGAGGCAAGCTGTATGGAGAGGCTCCACGACGATTAAG 272

RESULT 7
US-09-416-833-5
; Sequence 5, Application US/09416833
; Patent No. 6197521
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allgretti & Witcoff, Ltd.
; STREET: 10 S. Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/416,833
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/204,740
```

```

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 6197521nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-416-833-5

Query Match
Best Local Similarity 2.5%; Score 37.4; DB 4; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.032;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

QY 315 CCACCTGTTCACTATCCAGAGCTGGACCCAGCTTGACCCCTACTGCTGATGCTCA 374
DB 108 CCATCTGCTGGGCAAGCTAGGACGACCCCAAGAAACCGTGCTGATTTACGGGCA 167
QY 375 CTTTGATGTGGTGCT--GCCCTGAAGAGGCTGGAGGTCGCCCATTTCTGTGGTTG 432
DB 168 CCTGGAGCTGACAGCTGCGCCCTGGAGAGCGGGTGGAGACGACGACCCCTTACCTTGGTG 227
QY 433 GAGCGTATGGCGCTCATCTATGTCGGGGCACACTGACGACAAAG 477
DB 228 GAGCGGAGGCAAGCTGTATGGAGAGGCTCCACGACGATTAAG 272

RESULT 8
PCT-US95-02521-5
; Sequence 5, Application PC/TUS9502521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; NUMBER OF SEQUENCES: 13
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02521
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-02521-5

Query Match
Best Local Similarity 2.5%; Score 37.4; DB 5; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.032;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

QY 315 CCACCTGTTCACTATCCAGAGCTGGACCCAGCTTGACCCCTACTGCTGATGCTCA 374
DB 108 CCATCTGCTGGGCAAGCTAGGACGACCCCAAGAAACCGTGCTGATTTACGGGCA 167
QY 375 CTTTGATGTGGTGCT--GCCCTGAAGAGGCTGGAGGTCGCCCATTTCTGTGGTTG 432
DB 168 CCTGGAGCTGACAGCTGCGCCCTGGAGAGCGGGTGGAGACGACGACCCCTTACCTTGGTG 227
```


OY 433 GACGCGTANGCGCTCATCTATGTCGGGCGACACTGAGACAG 477
Db 228 GACGCGGAGGCAAGCTGTATGGAGAGGCTCCACGCGAGATAG 272

RESULT 9

US-09-120-663-1
; Sequence 1, Application US/09120663
; Patent No. 6228644
; GENERAL INFORMATION:
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMILOYORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,663
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,105
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1661
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-120-663-1

Query Match 2.4%; Score 36.6; DB 4; Length 5517;
Best Local Similarity 57.4%; Pred. No. 0.41;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
OY 235 TTGCGAAATACATCTTAAGTCTTCTACAGTGTGACGACCACTTTATCCAGCAT 294
Db 1000 TTGGTAAAGACACCCAGCACTATCTGGCGCAGCAGCAGCGTAGCCAGCAT 1059
OY 295 GAATCGTGAAGATATAGCCACTGTCTACTATCCAAAGGCTGGAGCCAGCT 349
Db 1060 CTGCTGTGACAAAGGCGACCTGTGTGATATCAAAAGCACCGCCAGCT 1114

RESULT 10
US-09-041-075A-24
; Sequence 24, Application US/09041075A
; Patent No. H002022
; GENERAL INFORMATION:
; APPLICANT: Heidler, Steven A
; APPLICANT: Raddling, Jeffrey A
; TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI

; FILE REFERENCE: X-11242 Sequence 1st
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 3168
; TYPE: RNA
; ORGANISM: Candida neoformans
; US-09-041-075A-24

Query Match 2.3%; Score 34.8; DB 1; Length 3168;
Best Local Similarity 35.4%; Pred. No. 1.1;
Matches 73; Conservative 26; Mismatches 107; Indels 0; Gaps 0;
OY 1074 CACAGTCAACTTCGCGATTCACCCCTGGACAGACATCCAAAGGTCTTGAAGTACGAA 1133
Db 1475 CCCACUCAAAGCUCGGGAGUACCCAUCCGCGGUCUACUCCGAGUACGUC 1534
OY 1134 GAACATTGTGGCTGATACAGAGTCCAGTTCATGTTGAGTGGCTTTGACCCCTGCC 1193
Db 1535 GCAGUUGUCUGCGCGCGACACCAUCCUUGCAGUCUACUCCUUCUCCGCCG 1594
OY 1194 CGTACGCCCTTGTGATGATGACAGGCTTGGCTACCAAGCTCTCCGACAGCCGTACAGTC 1253
Db 1595 UUDCAUCCUUCUCCGCGCGCGCGGCGGAUCCACGUCGCUCCUCCAGCCUUGAUGC 1654
OY 1254 CGTCTCCCGGAAGTAAATTTACTG 1279
Db 1655 CGUCCUACGUGGCCAACAUUUCG 1680

RESULT 11
US-09-041-075A-19
; Sequence 19, Application US/09041075A
; Patent No. H002022
; GENERAL INFORMATION:
; APPLICANT: Heidler, Steven A
; APPLICANT: Raddling, Jeffrey A
; TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence 1st
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 3220
; TYPE: DNA
; ORGANISM: Candida neoformans
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1888)..(1939)
; US-09-041-075A-19

Query Match 2.3%; Score 34.8; DB 1; Length 3220;
Best Local Similarity 48.1%; Pred. No. 1.1;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
OY 1074 CACAGTCAACTTCGCGATTCACCCCTGGACAGACATCCAAAGGTCTTGAAGTACGAA 1133
Db 1475 CCCACUCAAAGTGGGATACCCATGATGATCGGCGGTATCTTCCGATACGTC 1534
OY 1134 GAACATTGTGGCTGATACAGAGTCCAGTTCATGTTGAGTGGCTTTGACCCCTGCC 1193
Db 1535 GCAGTGTGCTGGCGCGCGACACCAUCCUUGCAGUCUACUCCUUCUCCGCCG 1594
OY 1194 CGTACGCCCTTGTGATGATGACAGGCTTGGCTACCAAGCTCTCCGACAGCCGTACAGTC 1253


```

1 EARLIER FILING DATE: 1997-05-23
2 EARLIER APPLICATION NUMBER: 60/047, 589
3 EARLIER FILING DATE: 1997-05-23
4 EARLIER APPLICATION NUMBER: 60/047, 593
5 EARLIER FILING DATE: 1997-05-23
6 EARLIER APPLICATION NUMBER: 60/047, 614
7 EARLIER FILING DATE: 1997-05-23
8 EARLIER APPLICATION NUMBER: 60/043, 578
9 EARLIER FILING DATE: 1997-04-11
10 EARLIER APPLICATION NUMBER: 60/043, 576
11 EARLIER FILING DATE: 1997-04-11
12 EARLIER APPLICATION NUMBER: 60/047, 501
13 EARLIER FILING DATE: 1997-05-23
14 EARLIER APPLICATION NUMBER: 60/043, 670
15 EARLIER FILING DATE: 1997-04-11
16 EARLIER APPLICATION NUMBER: 60/056, 632
17 EARLIER FILING DATE: 1997-08-22
18 EARLIER APPLICATION NUMBER: 60/056, 664
19 EARLIER FILING DATE: 1997-08-22
20 EARLIER APPLICATION NUMBER: 60/056, 876
21 EARLIER FILING DATE: 1997-08-22
22 EARLIER APPLICATION NUMBER: 60/056, 881
23 EARLIER FILING DATE: 1997-08-22
24 EARLIER APPLICATION NUMBER: 60/056, 909
25 EARLIER FILING DATE: 1997-08-22
26 EARLIER APPLICATION NUMBER: 60/056, 875
27 EARLIER FILING DATE: 1997-08-22
28 EARLIER APPLICATION NUMBER: 60/056, 862
29 EARLIER FILING DATE: 1997-08-22
30 EARLIER APPLICATION NUMBER: 60/056, 887
31 EARLIER FILING DATE: 1997-08-22
32 EARLIER APPLICATION NUMBER: 60/056, 908
33 EARLIER FILING DATE: 1997-08-22
34 EARLIER APPLICATION NUMBER: 60/048, 964
35 EARLIER FILING DATE: 1997-06-06
36 EARLIER APPLICATION NUMBER: 60/057, 650
37 EARLIER FILING DATE: 1997-09-05
38 EARLIER APPLICATION NUMBER: 60/056, 884
39 EARLIER FILING DATE: 1997-08-22
40 EARLIER APPLICATION NUMBER: 60/057, 669
41 EARLIER FILING DATE: 1997-09-05
42 EARLIER APPLICATION NUMBER: 60/049, 610
43 EARLIER FILING DATE: 1997-06-13
44 EARLIER APPLICATION NUMBER: 60/061, 060
45 EARLIER FILING DATE: 1997-10-02

```

Query Match	2.38;	Score 34.2;	DB 4;	Length 3542;
Best Local Similarity	49.28;	Pred. NO. 1.9;		
Matches	90;	Conservative	0;	Mismatches 93;
			Indels	0;
			Gaps	0;

QY	1501	AAA	1503	QY
Db	2657	GAA	2659	Db

Search completed: July 1, 2003, 07:48:24
Job time : 122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 07:07:34 ; Search time 269 Seconds

(without alignments)
8327.258 Million cell updates/sec

Title: US-10-014-896-1

Sequence: 1 atgctcagcggtgcgttg.....ctcaccgtcacacactgtga 1509

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	74	4.9	280	10 US-09-783-590-10310	Sequence 10310, A
2	46	3.0	1445	12 US-10-109-860-1	Sequence 1, App11
3	42	2.8	2710	9 US-10-037-270-931	Sequence 931, App
4	41.6	2.8	1988	9 US-10-073-885-54	Sequence 54, App
5	41.6	2.8	1997	10 US-09-925-301-555	Sequence 555, App
6	40.8	2.7	351	10 US-09-960-352-8484	Sequence 8484, App
7	40.4	2.7	2659	9 US-10-073-885-29	Sequence 29, App1
8	39.8	2.6	14536	9 US-10-160-758-6	Sequence 6, App11
9	39.8	2.6	14536	9 US-10-160-758-7	Sequence 7, App11
C 10	38.8	2.6	699	9 US-10-184-644-138	Sequence 138, App
C 11	38.8	2.6	699	9 US-10-184-644-138	Sequence 138, App
12	37.4	2.5	273	10 US-09-799-945-5	Sequence 5, App11
13	37.4	2.5	466	9 US-09-918-945-2716	Sequence 2716, App
14	36.8	2.4	298	9 US-10-108-605-280	Sequence 280, App
15	36.6	2.4	708	9 US-10-123-155-298	Sequence 298, App
16	36.6	2.4	5517	10 US-09-879-248-7	Sequence 7, App11
C 17	35.6	2.3	546	9 US-09-918-995-27597	Sequence 27597, A
C 18	35.4	2.3	1049	9 US-10-123-155-358	Sequence 358, App
19	35.2	2.3	734	9 US-10-184-644-458	Sequence 458, App

20	35.2	2.3	734	9	US-10-184-634-458	Sequence 458, App
21	35.2	2.3	1731	9	US-09-738-626-148	Sequence 148, App
22	35.2	2.3	3309400	9	US-09-738-626-1	Sequence 1, App11
23	35	2.3	2560	9	US-10-045-202-1	Sequence 1, App11
24	34.8	2.3	3168	9	US-09-742-580-24	Sequence 24, App1
25	34.8	2.3	3168	9	US-09-742-581-24	Sequence 24, App1
26	34.8	2.3	3168	10	US-09-742-582-24	Sequence 24, App1
27	34.8	2.3	3220	9	US-09-742-580-19	Sequence 19, App1
28	34.8	2.3	3220	9	US-09-742-581-19	Sequence 19, App1
29	34.8	2.3	3220	10	US-09-742-582-19	Sequence 19, App1
30	34.4	2.3	2065	9	US-10-122-805-5	Sequence 5, App11
31	34.2	2.3	1278	9	US-10-124-800-9	Sequence 9, App11
32	34.2	2.3	3542	9	US-09-809-391-63	Sequence 63, App1
C 33	34.2	2.3	8730	9	US-10-124-800-1	Sequence 1, App11
C 34	34	2.3	802	9	US-10-184-634-312	Sequence 312, App
C 35	34	2.3	802	9	US-10-184-634-312	Sequence 312, App
36	33.8	2.2	1344	9	US-10-153-668-413	Sequence 413, App
37	33.8	2.2	2673	9	US-10-152-668-301	Sequence 301, App
38	33.8	2.2	88421	9	US-09-976-059-1	Sequence 1, App11
39	33.6	2.2	1027	10	US-09-803-589-13	Sequence 13, App1
40	33.2	2.2	659158	9	US-09-771-208-20	Sequence 20, App1
41	33	2.2	1872	10	US-09-815-242-9401	Sequence 9401, App
42	33	2.2	2026	12	US-10-044-090-648	Sequence 648, App
C 43	32.8	2.2	3316	9	US-10-242-056-58	Sequence 58, App1
C 44	32.6	2.2	594	9	US-10-123-155-10	Sequence 10, App1
C 45	32.4	2.1	430	10	US-09-960-352-15018	Sequence 15018, A

ALIGNMENTS

RESULT 1
US-09-783-590-10310/C
Sequence 10310, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: DILLON, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16-2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 10310
LENGTH: 280
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (77)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (14)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (44)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (85)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (111)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (151)

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (182)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (203)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (228)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (244)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (246)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (259)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (278)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-10310
```

Query Match 4.9%; Score 74; DB 10; Length 280;
Best Local Similarity 96.1%; Pred. No. 1.3e-14;
Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 489 GCATTAAGGAGGCTTGGAGCTCTGATCAGAGAACTACATCCCGAAGATCTT 548
DB 77 GGCATTAAGGAGGCTTGGAGCTCTGATCAGAGAACTACATCCCGAAGATCTT 18
QY 549 CTTATTTCTCTGGGCC 565
DB 17 CTTATTTCTCTGGGCC 1
```

RESULT 2
US-10-109-860-1
Sequence 1, Application US/10109860
Patent No. US20020142421A1
GENERAL INFORMATION:
APPLICANT: SHAO, Wei et al.
TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
FILE REFERENCE: C001179DIV
CURRENT APPLICATION NUMBER: US/10/109,860
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/814,951
PRIOR FILING DATE: 2001-03-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1445
TYPE: DNA
ORGANISM: Homo sapiens
US-10-109-860-1

Query Match 3.0%; Score 46; DB 12; Length 1445;
Best Local Similarity 51.6%; Pred. No. 0.00017;
Matches 131; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

```
QY 326 CTAATCCAGGCTGGAGCCCGAGCTTGACGCCCTACCTGATGAGGCTCACTTTGATGTG 385
DB 202 CTACTATATGACACCAACCCCTACACTCTCTCTCATCTTGTCTCAACTCCACAGGATGTG 261
QY 386 TGGCTGCCCCGAGAGAGGCTGGAGGTGCCCATTTCTCTGGGTT---GGAAGCGTGAAG 442
DB 262 TGGCTGCTTCAAGAGAACTGAGATGACAGACCCCTTTGAGGCTTCAAGGATTTCTGAGG 321
QY 443 GCGTCACTATGATGCGGGGACACTGGAGACAGAACTCTGATGAGCATTAATGACAG 502
DB 322 GCTACATCTATGACAGGGGTGCCAGAGCATGAAGTGGCTGACGATCCGATCTGAGAG 381
```

```
QY 503 CTTGGAGGCTCCGCTGATCAGAGAACTACATCCCGAAGATCTTTCTCATTTCTGCG 562
DB 382 CTGTGAGAGGCTGGAAGGTGAGGGCCACCGGTTCCCGAAGCAATCCAGATGACCTTG 441
QY 563 GCCATGATGAGGAG 576
DB 442 TGGCTGATGAGGAG 455
```

RESULT 3
US-10-037-270-931
Sequence 931, Application US/10037270
Publication No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunhui
APPLICANT: Wang, Zhilwei
APPLICANT: Tillinghast, John
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pL_genes Version 1.0
SEQ ID NO 931
LENGTH: 2710
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (218)..(1645)
US-10-037-270-931

Query Match 2.8%; Score 42; DB 9; Length 2710;
Best Local Similarity 54.6%; Pred. No. 0.0062;
Matches 106; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

```
QY 319 CTGTTCACTATCCAGGCTCGAGCCCAAGCTTGGAGCCCTACCTGATGAGGCTCACTTT 378
DB 458 CTGCTCGGAGGCTGGCTGTGACCAAGAAAGACCGTGTGATTAAGGACACTG 517
QY 379 GATGTGTGCT---GCCCTGAAGAGAGGCTGGAGGTGCCCATTTCTCTGGGTTGAG 435
DB 518 GATGTGAGCTGACAGCCCTGAGAGCGCTGGGACAGCGAGCCCTTACCCCTGTGGAG 577
QY 436 CGTATGCGCATATGATGATGCTGGGGCACTGAGACGACAAAGAACTCTGTGATGAGCTTA 495
DB 578 CGAGACGCAAGCTGTATGAGAGAGTTGCACTGATATTAAGGGCCGCGTGGCGGCTG 637
QY 496 CTGAGGCTTGA 509
DB 638 ATAAAGCCCTGA 651
```



```

GENERAL INFORMATION:
APPLICANT: Gudkov, Andrei
            Kazarov, Alexander
            Mazo, Il'ya
            Roninson, Igor B
TITLE OF INVENTION: Methods for Identifying Genetic
                    Suppressor Elements and Genes Associated with Malignant
                    Growth in Cancer Cells
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Allgretti & Witcoff, Ltd.
STREET: 10 S. Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,946
FILING DATE: 06-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/204,740
FILING DATE: 04-Mar-1994
ATTORNEY/AGENT INFORMATION:
NAME: No. US2002009028Alman, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,354-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-799-946-5

Query Match      2.5%; Score 37.4; DB 10; Length 273;
Best Local Similarity 55.8%; Pred. No. 0.05;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

OY 315 CCACCTGTCACATCCAGGCGGAGCCCGGAGCCCTGACCTGCTGATGCGCTCA 374
DB 108 CCAATGCTGCGGAGAGTGAAGCGAGCCCGGAGGAAACCGTGTGCTTACCGGCA 167
OY 375 CTTTGATGTGCTGCT--GCCCTGAAGAGGCTGGAGGAGGCCCATCTCTGGGTG 432
DB 168 CCGGAGAGTGCACGCTCGCGCTGGAGAGAGGCTGGAGAGCGGAGCGGCTTACCTTGGTG 227
OY 433 GAGCGGATGGCTCATATGTCGCGGAGGACACTGAGGACAGCAAG 477
DB 228 GAGCGGAGAGGCAAGCTGTATGGAGAGGCTCCAGGAGGAGATGAG 272

RESULT 13
US-09-918-995-2716
; Sequence 2716, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995

```

```

CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2716
LENGTH: 466
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(466)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2716

Query Match      2.5%; Score 37.4; DB 9; Length 466;
Best Local Similarity 52.2%; Pred. No. 0.072;
Matches 108; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

OY 318 CCGTTCACATATCAAGGCTCGAGCCCGGAGCTTGAGCCCTACCTGATGGCTCACTT 377
DB 250 CCGTTCACATATCAAGGCTCGAGCCCGGAGGACCAACCTTACCTTCTCACTTCTCACTTCCAC 309
OY 378 TGATGTGTGCTGCTGCCCCGAGAGGCTGGAGGAGTGCCTCATCTCTGGGT--GGA 434
DB 310 GGATGTGTGCTGCTGCTTCAAGGAACATTTGAGTACAGCCCTTTGAGGCTTCAAGA 369
OY 435 GCGTATGGCGTCATCATATGTCGCGGAGGACACTGACGACGAAGAAGCTGTGATGGCAAT 494
DB 370 TTCTGAGGCTCATATGATGCGAGGGGTCGACGACATGAATGCTGACATCCACTA 429
OY 495 ACTGAGGCTTGAAGCTCTGCTGAT 521
DB 430 CCGGAGCTGTGAGAGGCTGAGGT 456

RESULT 14
US-10-108-605-280
; Sequence 280, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Steam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patent In Ver. 2.1
; LENGTH: 298
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-280

Query Match      2.4%; Score 36.8; DB 9; Length 298;
Best Local Similarity 61.5%; Pred. No. 0.086;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 1177 GCCTTGACCCCTCCGCTGAGCCCTTGTATACAGGCTTGGCTACAGCTGCTC 1236
DB 139 GCGTGTCTTCTCTGCGCGCTGCTTCCGCGAGAGAGAGGATGCTGCTGCTC 198
OY 1237 GCGGAGCGGTACAGTCCGCTTCCGGAAGTCAAT 1272
DB 199 GACCAAGGCGCGGACATCAACAGCCCAAGCTGAT 234

```

RESULT 15
US-10-123-155-298 Application US/10123155
; Sequence 298, Publication NO. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geriltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 298
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-298

Query Match 2.4%; Score 36.6; DB 9; Length 708;
Best Local Similarity 6.7%; Pred. No. 0.18;
Matches 37; Conservative 142; Mismatches 372; Indels 0; Gaps 0;

QY 78 GGGCCGAGAGGGGAGAGATCAAGAGGCGTGGCAATCCCTTCAGTTGAGCAAGA 137
DB 112 VKMPOLLSTYLEENKLTPEKLSLSMLQELYNHMLSTISPAFTGLHNLHL 171
QY 138 GGAACCGGTCGCGATGAAGAGGCGTGAAGGTCGATCCAGATTCACAGTGA 197
DB 172 NSNRLOMINSKWDALPNLEIIMIGENPIIRIDMNFPLINRSLVIAGINLTETPDNA 231
QY 198 TAGCTGTGAGAGTCAATACATACAGAGGCTGCTGAGTTCGAAATATCATTAAGT 257
DB 232 LVGLLENLSTISFYDNRLIKVPHVALQKVMFLDLINKPILNIRIGRDFSNMLHLKELGI 291
QY 258 CTTTCCAGAGTGTGAGCAGCAGCTTATCCAGCATGAAGTGTGGAAGAGTATAGCCA 317
DB 292 NNPPELISIDSLAVNDLPDIRKTEATNPNRLSTIHPNAFFRLPLESLMLNSALSLYH 351
QY 318 CCGTTCATATCAAGAGTCCGAGCCAGCTGCAGCCCTGATGCTGATGCTGCTGCT 377
DB 352 GTIESLPNLKEISIHSPRICDCYIRMMNKNKTINIRMEPDSLFCVDPPEFOGQANROYH 411
QY 378 TGAATGAGTCCCTGCGCCCTGAGAGGCTGGAGGTGCCCCATTCCTGAGTTGAGCG 437
DB 412 FRDMEICLPLIAPESPSNUNVAGSYVSFHCRAFAEPQELIYWTIPSGOKLLPWTMD 471
QY 438 TGAATGAGTCCATATGATGAGGCGGACACACAGAGAGTGTGATGATGATGATGAT 497
DB 472 KFYVHSEGLDINGVPEGLTYTATNLVAGDLKSVMLKVDGSEFPDNNGLNIRKIR 531
QY 498 GCAGGCGTGTGAGTCCCTGCTGATGAGAGTATACATCCCGGAGAGATCTTTCATTTTC 557
DB 532 IQANSVLYSWKASKILKSSVKWTATATVKTENSHADGASARIPSDVKYINLTLPSTERYK 591

QY 558 TCTGGCCATGATGAGAGTATCATCAGGAGGAGGCGCTCAGAGATCTCAGCCCTGCTACA 617
DB 592 CIDIPTIYOKNRKRCVNTTFTGLHPDQREYERKNTTTLMACTGGLGIIIVTCLISCLSP 651
QY 618 GTCAGGGGCG 628
DB 652 EKNCDGHSYV 662

Search completed: July 1, 2003, 09:22:14
Job time : 283 secs

QY 1381 AACGATCCATGATGATCAACGAGAAATCTCAGTCCAGACCTATGAGACCCCAAGTGANA 1440
DB 432 LysArgIleHisGlyValAsnGluLysIleSerValGlnAlaTyrGluThrGlnValLys 451
QY 1441 TTCATCTTGGTGGTATTCAGAAATGCTGACACAGACGAGGACGACCTTCTCAGCTGCAC 1500
DB 452 PheIlePheGluLeuIleGlnAsnAlaAspThrAspGlnGluProValSerHisLeuHis 471
QY 1501 AAACGTG 1506
DB 472 LysLeu 473
RESULT 2
AAB40345
ID AAB40345 standard; Protein; 166 AA.
AC AAB40345;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF109 polypeptide sequence SEQ ID NO:218.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW valineary; antipsoptic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000WO-0508621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX
PI Shinkets RA, Leach M;
DR WPI; 2000-602362/57.
XX
XX N-PSDB; AAC74554.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX
XX Claim 11; Page 547; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; valineary;
XX antipsoptic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antihypertensive; and antianemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy.
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX
SQ Sequence 166 AA:
XX
XX
Alignment Scores:
Pred. No.: 4,41e-71 Length: 166
Score: 827.00 Matches: 164
Percent Similarity: 99.398 Conservative: 0
Best Local Similarity: 99.398 Mismatches: 1
Query Match: 29.944 Indels: 0
DB: 21 Gaps: 0
US-10-014-896-1 (1-1509) x AAB40345 (1-166)
QY 142 CCGTCGCGATGAAAGAGCGCTGAAAGGCTCCATCCAGATTCCAAAGTGACTTTAGC 201
DB 2 ArgValAlaMetLysGluAlaLeuLysGlyAlaIleGlnIleProThrValThrPheSer 21
QY 202 TCGAGAAAGTCCAAATCTCAGCCCGCTGGTGGTGGGAAATATCATCATTAAGCTTT 261
DB 22 SerGluLysSerAsnThrThrAlaLeuAlaGluPheGlyLysTyrIleHisLysValPhe 41
QY 262 CCTACAGTGTGACAGCAGCAGCTTTATTCAGATGAGTGTGGAAGATATGACCACTG 321
DB 42 ProThrValValSerThrSerPheIleGlnHisGluValValGluGluTyrSerHisLeu 61
QY 322 TTCACATTCGAAAGCTCGGAGACCCAGCTTGACAGCTTACTGCTATGGCTACTTTGAT 381
DB 62 PheThrIleGlnLysSerAspProSerLeuGlnProTyrLeuMetAlaHisPheAsp 81
QY 382 GTGTCGTCGCCCTGAAAGAGGCTGGAGTGCCCATTTCTGTGGGTGGAGCTGAT 441
DB 82 ValValProAlaProIleGluGluGlyTyrPheValProProPheSerGlyLeuLysAsp 101
QY 442 GCGCATCATGTATGCTGGGACACATGACGACAGCAAGACTCTGTGATGGCATTTACTG 501
DB 102 GlyValIleTyrGlyTyrPheLysAspPheLysAsnSerValMetAlaLeuGln 121
QY 502 GCGTTGGAGCTCTGCTGATCAGGAAGTACATCCCGAAGATCTTTCTTCAATTCTCTG 561
DB 122 AlaLeuGluLeuLeuLeuIleArgLysTyrIleProArgArgSerPhePheIleSerLeu 141
QY 562 GCGCATGATGAGAGTCAATCAGGAGGACAGGAGCTCAGAGATCTCAGCCCTGATACGTA 621
DB 142 GlyHisAspGluLeuSerSerGlyThrGlyAlaGlnArgIleSerAlaLeuGlnSer 161
QY 622 AGGGCGCTCCAGCTA 636
DB 162 ArgGlyValGlnLeu 166
RESULT 3
AAO04998
ID AAO04998 standard; Protein; 120 AA.
XX
XX AAO04998;
AC
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 18890.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX Homo sapiens.
 XX OS
 XX PN MO200164835-A2.
 XX PD 07-SEP-2001.
 XX PE 26-FEB-2001; 2001MO-US04927.
 XX PR 28-FEB-2000; 2000US-0515126.
 XX PR 18-MAY-2000; 2000US-0577409.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-514838/56.
 DR N-PSDB: AAI84929.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20: SEQ ID NO 18890; 1399pp + Sequence Listing; English.
 XX
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 120 AA:
 SQ
 Alignment Scores:
 Pred. No.: 1,39e-45 Length: 120
 Score: 562.50 Matches: 114
 Percent Similarity: 79.72% Conservative: 0
 Best Local Similarity: 79.72% Mismatches: 0
 Query Match: 20.37% Indels: 29
 DB: 22 Gaps: 1
 US-10-014-896-1 (1-1509) x AAO04998 (1-120)
 OY 1 ATGGCTACAGCGGTGCTTGGCTGCGTGGTGGTATGCTGCTAGTTTCCCT 60
 DB 7 MetAlaGlnArgCysValcLysValIleuAlaIleuValAlaMetIleuLeuValPhePro 26
 OY 61 ACCGTTCAGATTCGATGGGCGGAGAGCGGGGAGCATCAAGGCGGTCCGGAATCCCT 120
 DB 27 ThrValSerArgSerMetGlyProArgSerGlyGlnHisGlnArgAlaSerArgIlePro 46
 OY 121 TCTCAGTTACAGCAAGAGGAGGAGCGTGGCAGGAAGAGGCCCTGAAAGGTTCCATCCAG 180
 DB 47 SerGlnPheSerLysGlnGlnIleArgValAlaMetLysGlnAlaIleu----- 61
 OY 181 ATTCCAACAGTACTTTAGCTCTGAGAAGTCCAACTACTACAGCCCTGGTGAGTTGGGA 240
 DB 61 ----- 61
 OY 241 AAATCATTCATAAAGCTTTCTCTACAGTGTGTCACACACAGTTTATCCAGATGAACTC 300
 DB 62 -----LysValPheProThrValValSerThrSerPheIleGlnHisGlnVal 77

OY 301 GTGGAAGAGTATAGCACACCTGTTCACTATCCAAAGGCTGGAGCCCAAGCTTGACGCCCTAC 360
 DB 78 ValGlnGlnTrpSerHisIleuPheThrIleGlnGlySerAspProSerLeuGlnProGly 97
 OY 361 CTGCTGATGAGCTCACTTTGATGTGGTGGCTGCGCCCTGAAGAGGCTGGAGGTGCCCCCA 420
 DB 98 LeuLeuMetAlaHisPheAspValValProAlaProGlnGlnGlyTrpGlnValProPro 117
 OY 421 TTCTCTGGG 429
 DB 118 PheSerGly 120
 RESULT 4
 ABP09238
 ID ABP09238 standard; protein; 116 AA.
 XX
 AC ABP09238;
 XX
 DT 25-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:18458.
 XX
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 PN MO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001MO-US10836.
 XX
 PR 30-MAY-2000; 2000US-206132P.
 PR 29-AUG-2000; 2000US-228716P.
 XX
 PA (CDRA-) CORAGEN CORP.
 XX
 PI Shinkets RA, Leach MD;
 XX
 DR WPI: 2002-106308/14.
 DR N-PSDB: ABN24990.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX
 PS Disclosure: SEQ ID 18458; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification)). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also

CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).

XX Sequence 116 AA;

Alignment Scores:

Pred. No.:	7.68e-37	Length:	116
Score:	472.00	Matches:	89
Percent Similarity:	90.43%	Conservative:	15
Best local Similarity:	77.39%	Mismatches:	11
Query Match:	17.09%	Indels:	0
DB:	23	Gaps:	0

US-10-014-896-1 (1-1509) x ABP09238 (1-116)

OY 178 CAGATTCACACAGTACTTACGCTGAGAGTCAATACAGCCCTGGCTGAGTTC 237
DB 2 Arg1LeuProthValSerPheSerHisGluGluSerHisThrAlaLeuAlaGluPhe 21
OY 238 GGAATAATACATTAATAAGTCTTCTACAGTGTGACACACACCTTATCCACATGAA 297
DB 22 GlyIuTyrlIleAgluYsAlaPheProthValPheHisSerHisLeuIleGlnHisGlu 41
OY 298 GTCTGGAGAGGTATACACACCTGTGATCCAGAGGTGGAGCCACAGCTTTCACAGCC 357
DB 42 ValValGlyIuYsIleSerHisLeuLeuThrValArgIleSerHisProSerLeuGlnPro 61
OY 358 TACCTGCTGATGCTCACTTGTATGTGTGCTGCTCCCTGAAGAAGCTGGAGTCC 417
DB 62 TyrMetLeuMetAlaHisPheAspValValProAlaSerGluGluGlyTTPGluValPro 81
OY 418 CCATTCCTGCTGGTGGAGCCGATGCGCTCATCTATGCTGGGGGACACATGGAGACAAG 477
DB 82 ProPheSerGlyLeuGluGlnAsnGlyPheIleHisGlyAlaGlyAlaLeuAspAsnLys 101
OY 478 AACCTGTGATGATGATACCTACAGCCCTGGAGCTGCTGCTGATC 522
DB 102 AsnSerValMetAlaValLeuGlnAlaLeuGlnLeuLeuLeuLeu 116

RESULT 5

ID AAY11540 standard; Protein: 93 AA.

XX AAY11540;

XX 16-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO:192.

KM Human; secreted protein; EST; expressed sequence tag; diagnosis;
KM forensic; gene therapy; chromosome mapping; signal peptide;
KM upstream regulatory sequence; cytokine activity; cell proliferation;
KM differentiation; haematopoiesis regulation; tissue growth regulation;
KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KM thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.
XX
XX W09906439-A2.
XX
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98WO-1B01233.
XX
XX 01-AUG-1997; 97US-0904468.
XX
XX
XX PA (GEST) GENSET.

XX Ducleert A, Dumas Milne Edwards J, Lacroix B;
PI WPI: 1999-153700/13.
DR N-PDB: AAX40258.

XX New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from liver, lung, large intestine, colon,
PT thyroid and pancreas tissue

PS Claim 27; Page 312; 398pp; English.

XX AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY11533 to
CC AAY11679, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 93 AA;

Alignment Scores:

Pred. No.:	7.3e-35	Length:	93
Score:	451.00	Matches:	93
Percent Similarity:	98.94%	Conservative:	0
Best local Similarity:	98.94%	Mismatches:	0
Query Match:	16.33%	Indels:	1
DB:	20	Gaps:	0

US-10-014-896-1 (1-1509) x AAY11540 (1-93)

OY 1 ATGGCTCAGCGGCGGCTTGGCTGCGGCTGGTATGCTGCTAGTTTCCT 60
DB 1 MetAlaGlnArgGlyValGlyValLeuAlaLeuValAlaMetLeuLeuLeuValPhePro 20
OY 61 ACCGCTCCAGATGATGGGCCCGAGAGCGGGAGCATCAAGGGCTGCGGAATCCCT 120
DB 21 ThrValSerArgSerMetGlyProArgSerGlyGlnHisGlnArgAlaSerArgIlePro 40
OY 121 TCTCAGTTCAAGCAAGAGACGCGCTGCGATGAAGAGCGCTGAAGGTCATCCAG 180
DB 41 SerGlnPheSerIysGluGlnArgValAlaMetGlyGlnAlaLeuLysGlyAlaIleGln 60
OY 181 ATTCCACACAGTACTTACCTGAGAGTCCAAATACACAGCCCTGGCTGAGTTCGGA 240
DB 61 IleProThrValThrPheSerSerGluLysSerHisThrThrAlaLeuAlaGluPheGly 80
OY 241 AAATACATTCATAAAGTCTTCTACAGTGTGTCAGACCA 280
DB 81 -AsnThrPheIleLysSerPheLeuGlnTrpSerAlaPro 93

RESULT 6

ID AAG92859 standard; Protein: 441 AA.

XX AAG92859;

XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 6613.
XX
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

QY 1405 -----AAATCTCACTCCAGCCTATGAGACCAAGTGAAA 1440
DB 422 LeuGIuAspLeuGIuLeuThrValArgIleTyrAspSerValValArg 437
RESULT 7
ID AAB79123 standard; Protein; 422 AA.
XX AAB79123;
AC AAB79123;
XX 30-APR-2001 (first entry)
DT
XX Corynebacterium glutamicum HA protein sequence SEQ ID NO:202.
DE
XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW fine chemical production; organic acid; proteogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diptheriae; genetic engineering;
KW Brevibacterium; environmental condition.
XX Corynebacterium glutamicum.
OS
XX W0200100842-A2.
PN
XX 04-JAN-2001.
PD
XX 23-JUN-2000; 2000MO-IB00911.
PE
XX 25-JUN-1999; 9905-0141031.
XX 08-JUL-1999; 99DE-1031636.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032126.
XX 09-JUL-1999; 99DE-1032127.
XX 09-JUL-1999; 99DE-1032128.
XX 09-JUL-1999; 99DE-1032129.
XX 09-JUL-1999; 99DE-1032226.
XX 14-JUL-1999; 99DE-1032920.
XX 14-JUL-1999; 99DE-1032922.
XX 14-JUL-1999; 99DE-1032924.
XX 14-JUL-1999; 99DE-1032928.
XX 14-JUL-1999; 99DE-1032930.
XX 14-JUL-1999; 99DE-1032933.
XX 14-JUL-1999; 99DE-1032935.
XX 14-JUL-1999; 99DE-1032973.
XX 14-JUL-1999; 99DE-1033002.
XX 14-JUL-1999; 99DE-1033003.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041390.
XX 03-SEP-1999; 99DE-1041391.
XX 03-SEP-1999; 99DE-1042088.
XX (BADI) BASF AG.
XX
XX Pompeius M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX WPI, 2001-061974/07.
XX N-PSDB; AAF71238.
XX
XX New isolated Corynebacterium glutamicum nucleic acid for production or
XX modulation of production of fine chemicals such as amino acids,
XX nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
XX or enzymes -
XX
XX Claim 20; Page 413-414; 712pp; English.
XX AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
XX and adaptation (HA) proteins given in AAB79023 to AAB79242. The

CC C. glutamicum HA genes (1) can be used in vectors for expression in host
CC cells and production of fine chemicals, such as, an organic acid,
CC proteogenic or nonproteogenic amino acid (preferred), purine or
CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC be modulated. The presence of (1) or HA proteins encoded by then are
CC used for diagnosing the presence or activity of Corynebacterium
CC diptheriae. (1) can be used to map the C. glutamicum genome or can be
CC used as markers for genetically engineered Corynebacterium or
CC Brevibacterium. The HA proteins encoded by the (1) are used to maintain
CC homeostasis in C. glutamicum or help the microorganism to adapt to
CC different environmental conditions.

CC Sequence 422 AA;

CC Alignment Scores:

CC Pred. No.: 2,16e-17 Length: 422
CC Score: 274.00 Matches: 106
CC Percent Similarity: 42.55% Conservative: 71
CC Best Local Similarity: 25.48% Mismatches: 161
CC Query Match: 5.92% Indels: 78
CC DB: 22 Gaps: 20

US-10-014-896-1 (1-1509) x AAB79123 (1-422)

QY 313 AGCCACCTGTCACATTCACAGGCTCGGACCCAGCTTCGACCCCTACCTCGATGGCT 372
DB 41 SerIleIleValThrValProGlySerAspProAspAlaGluProLeuThrLeuGly 60
QY 373 CACTTGTATGTGTGCTCCCTCCAGAGGCTGGAGGCTCCCATCTCTGGGTTG 432
DB 61 HisThrAspValValProValAspLeuProLysTrpThrLysAspProGlyAlaGlu 80
QY 433 GAGCGTGATGGCGTCATCTATGCTGGCGCACCTGGACGCAAGAACTCTGTATGCA 492
DB 81 IleSerAspGlyGluIleTrpGlyArgIleValAsp-----MetLeu 95
QY 493 TTACGTGACGCGCTTGAGCTCTGCTGATGACGAATGATCCCGAAGATCT----- 545
DB 96 PheIleThrAlaThrGlnAlaAlaValThrArgin---ValAlaArgGluGlyLeu 114
QY 547 -----TTCTCATTTCTCTGGCCATGATGAGGATCATCGAGGAGGCT 594
DB 115 ArgGlyThrLeuThrPheValGlyValAlaAspGluGluAlaArgGlyLeuGlyAla 134
QY 595 CAGAGAGTCTCA-----GCCCTGCTACAGCA 621
DB 135 LysTrpLeuSerGluGluHisGlnAsnLeuPheSerTrpLysAsnCysLeuSerGluSer 154
QY 622 AGCGGCGTCACGCTGCTCATTTGTCGAGGAGGCGGCTTCATCTGATGATTCATT 681
DB 155 GlyGlySerHisLeuPro-----ValHisAspGly----- 164
QY 682 CCTAAGTTCAGAACCCATCGCTGATGTCAGATGTCAGAGAGGTTCCATGAACCTC 741
DB 165 -----SerAspAlaValValIleAsnValGlyGlyAlaAlaGlnArg 180
QY 742 ATGCTGCAGTAACATGATGCTTCAGGCCATCTTACGCTCC---CCAAGAGAGCAAGC 798
DB 181 ArgIleHisValAsnGlyAspAlaGlyHisGlySerIleProPheAspTrgAspSerAla 200
QY 799 ATTGCGATCTTCAGAGCTGCTGTCAGCCGATTCGAGAGCAAGCAATGCTATCACTA--- 855
DB 201 IleValIleGlyGlyValAlaArgAlaGlyIleAlaAlaAspLeuLysValAlaLys 220
QY 856 -----TTTGAAGGGGAGCAGTGTGATGCTATTGACAGCAAGTGGCAATGATTT--- 906
DB 221 AspAspIleTrpGlnGlyPheVal-----GlnAlaHisArgPheAsp 234

```

QY 907 CCCTCCCTGTAATTAATCTGAGC-----AACCCAGGCTATTGTGACCACTT--- 957
Db 235 ProGluThrGluGlnAlaLeuSerGlyThrSerProGluAlaTyrAlaGluPheGly 254
QY 958 ---ATAAGCAGGTTTATGAGAGAGAAATCCCTTAACCAATCATATACAGACCAACG 1014
Db 255 GlyLeuSerThrPhe-----AlaHisAlaValSerHisLeuThrIle 268
QY 1015 GCACATCAATTAATCAAGCAGGAGGTCAGTTCATGTCATCCCGCAGTGGCCAGGCC 1074
Db 269 AlaGlnThrValValArgAlaGlyGlnAlaIleAsnValLeuProSerHisAlaTyrLeu 288
QY 1075 ACAATCACTCCGAGATTCACCCCTGGACACACA-----GTCCAGAGGTCCTAGAA 1125
Db 289 GluLeuAspIleArgThrLeuProGlyGlnThrAsnAspTyrValAspAspThrLeuArg 308
QY 1126 CTCACGAAAGAAATGTCGCTGATTAACAGAGTCACAGTCCAGTCCAGTCCGCTTGAC 1185
Db 309 AlaAlaLeuGlyAspLeuAlaAspGluValGluIleGlnHisLeuIleSer-----Glu 326
QY 1186 CCCCTCCCGCTCAGCCCTTCTGTGACACAGCCCTTGCGTACAGCTCCGCGCAGACC 1245
Db 327 GluAlaThrValSerProThrAspSerArgLeu-----TyrAsnThrLeuGluVal 344
QY 1246 GTACAGTCCGCTCTCCCGAAGACATATTAATCTGACCCCGACTTACTCTATGTGCAACACA 1305
Db 345 LeuGlyAspRhePheProAspAlaProVal---ValIProIleIleSerGlyGlySer 363
QY 1306 GACAGCCGATCTTTTACAAATCAACACACACGACATGACAGTTCACCCCATCTCAATA 1365
Db 364 AspLeuArgPheGlyArgArgLeu---GlyGlyValGlyTyrGlyPheAlaValHisAla 382
QY 1366 CAG-----CCTGAAGACTTCAAAGCATCCAGTCCAGTCAACAGC----- 1404
Db 383 ArgGluArgThrLeuAlaGlnIleMetGlyGlnLeuHisSerHisAspGluAlaLeuTyr 402
QY 1405 -----AAATCTCAGTCCAAAGCCTATGACACCCCAAGTGA 1440
Db 403 LeuGlnAspLeuGlnLeuThrValArgGlyTyrAspSerValValArg 418

RESULT 8
AAB79819 standard; Protein; 422 AA.
XX
AC AAB79819;
XX
DY 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:372.
XX
KW fine chemical production; metabolic pathway protein; MP protein;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN W0200100843-A2.
PD
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-IB00923.
XX
PR 25-JUN-1999; 9905-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 9905-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031426.

```

```

PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 08-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032300.
PR 09-JUL-1999; 99DE-1032322.
PR 09-JUL-1999; 99DE-1032326.
PR 09-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 9905-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 31-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.

(BADI ) BASF AG.
XX
PA Pompejus M, Krieger B, Schroeder H, Zelder O, Habernauer G;
PI WPI: 2001-137957/14.
XX
DR N-PSDB; AAF71938.
XX
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX pathway proteins, useful for producing fine chemicals in
XX microorganisms, including organic acids, nonproteinogenic amino acids,
XX and purine and pyrimidine bases -
XX
XX Claim 20; Page 717-718; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX MP nucleic acids are useful for the production of fine chemicals
XX in microorganisms, including organic acids, nonproteinogenic amino
XX acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 422 AA;

```

Alignment Scores:

Pred. No.:	2,16e-17	Length:	422
Score:	274.00	Matches:	106
Percent Similarity:	42.55%	Conservative:	71
Best Local Similarity:	25.48%	Mismatches:	161
Query Match:	9.92%	Indels:	78
	22	Gaps:	20

US-10-014-896-1 (1-1509) x AAB79819 (1-422)

```

OY 313 AGCCACCTGTTCACTTCCAGGCGTGGAGCCCAAGCTTGACCCCTACCTGCTGATGCT 372
    |||  ::  |||::|  |||::|  |||::|  |||::|  |||::|  |||::|  |||::|
Db 41 SerIleIleValThrValProGlySerAspProAspAlaGluProIleuThrLeuGly 60
OY 373 CACTTGTATGGTGGCTGCCCGTGAAGAAGGCTGGAGAGTGGCCCATCTGCGGTG 432
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 61 HisThrAspValValProValAspLeuProLysThrIleLysAspProIleGlyAlaGlu 80
OY 433 GAGCGTGTATGCGCTCATGTATGCTCGGGGACACCTGGACACAGCAAGACTGTGATGCA 492
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 81 IleSerAspGlyGlnIleTyrGlyArgGlySerValAsp-----MetLeu 95
OY 493 TTACTGAGCGCTTGGAGCTGCTGCTGATCAGAGAGTACATCCCGGAGATCT----- 546
    ::  ||  ::  ||  ::  ||  ::  ||  ::  ||  ::  ||  ::  ||  ::  ||
Db 96 PheIleThrAlaThrGlnAlaAlaValThrArgGln---ValAlaArgGluGlyLeu 114
OY 547 -----TTCTCATTTCTCTGGCCATGATGATGAGAGTCAATGACGAGGAGGCT 594
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 115 ArgGlyThrIleuThrPheValGlyValAlaAlaAspGluAlaArgGlyLeuGlyAla 134
OY 595 CAGAGATCTCA-----GCGCTGCTACACTCA 621
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 135 LysThrLeuSerGluGlnIleAsnIleuPheSerTyrLysAsnGlySerGluSer 154
OY 622 AGGGCGTCCGACGCTTCATGCTTGAGAGGAGGCGGCTTCATCTTGATGATTTGAT 681
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 155 GlyIleSerHisLeuPro-----ValHisAspGly----- 164
OY 682 CCTAACTTCAAGAGCCCATCGCTTATGTCAGTCTCAGAGAGGTTCCATGACCTC 741
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 165 -----SerAspAlaValValIleAsnValGlyGluLysGlyAlaIleAlaArg 180
OY 742 ATGTGCAATAAACATGACTTTCAGGCCACTCTTCACCTCT---CCAAAGAGACAGC 798
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 181 ArgIleHisValIleAsnGlyAspAlaGlyHisGlySerIlePheAspArgAspSerAla 200
OY 799 ATTGGCATCCCTTGGACGCTGCTGACGCCGATTTGGAGACACACCAATGCCATGATA-- 855
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 201 IleValLysIleGlyGluValAlaAlaArgArgIleAlaAlaAlaAspLeuLysValAlaLys 220
OY 856 -----TTTGGAGCGGAGACAGTGTGATGCTATTTGACAGCAATGGCAATGACTT-- 906
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 221 AspAspIleTyrPheGlnGlyPheVal-----GlnAlaHisArgPheAsp 234
OY 907 CCCCTCCCTGTCATATATATCCCTGAGC-----AACCCATGGCTATTTGAACACCT-- 957
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 235 ProGluThrGluGlnAlaLeuIleuSerGlyThrSerProGluAlaTyrAlaGluPheGly 254
OY 958 ---ATAAGCAGGTATATGAGAGAAATCCCTTAACCAATGCAATATAGAGACACACAGC 1014
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 255 GlyLeuSerArgPhe-----AlaHisAlaValSerHisLeuThrIle 268
OY 1015 GCACCTCAACATATTCMAAGCAGGGGTCAAGTTCATATGATCCCGCCAGTGGCCAGGCC 1074
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 269 AlaGlnThrValValArgAlaGlyGlnAlaIleAsnValLeuProSerHisAlaTyrLeu 288
OY 1075 ACAGTCACTCCCGATTCACCCCTGGAGACACA-----GTCCAAAGAGCTCTGAGAA 1125
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 289 GluLeuAspIleArgThrLeuProLysGlnThrAsnAspTyrValAspAspThrLeuArg 308
OY 1126 CTCACGAAAGACATTTGGTGTGATTAACAGAGTCCAGTTCATGCTGTAGTGCCTTTGAC 1185
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db 309 AlaAlaLeuGlyAspLeuAlaAspGluValGluIleGlnHisLeuIleSer-----Glu 325
OY 1186 CCCCTCCCGTACGCGCTTGTGATGACAGGCGCTTGGGCTACAGCTCTCGCCAGAC 1245
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 327 GluAlaThrValSerProThrAspSerArgLeu-----TyrAsnThrLeuGlnLysVal 344
OY 1246 GTACACTCCGCTTTCCCGGAGCTCAATATTTACTGCCCGCAATCTTCTATTGGCAACACA 1305
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 345 LeuGlyAspPhePheProAspAlaProVal---ValProIleIleSerSerGlyLysSer 363
OY 1306 GACAGCCGATCTTTACAAACCTGCACCACTGGCATCTACAGTGTCTACCCCATACADA 1365
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 364 AspLeuArgPheGlyArgArgLeu---GlyGlyValGlyTyrGlyPheAlaHisAla 382
OY 1366 CAG-----CCTGAGACTTTCMAAGCATCATGAGAGTCAAG----- 1404
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 383 ArgGluArgThrLeuAlaGlnLametGlyGlnLeuHisSerHisAspGluLysLeuTyr 402
OY 1405 -----AAATCTCAGTCCAGGCTTATGAGACCCAAAGTGA 1440
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 403 LeuGluAspLeuGluLeuThrValArgGlyTyrAspSerValValArg 418

```

RESULT 9

AAU41311 standard; Protein; 458 AA.

AAU41311;

13-FEB-2002 (first entry)

Propionibacterium acnes immunogenic protein #2207.

SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 dermatological; osteopathic; neuroprotectant.

Propionibacterium acnes.

W0200181581-A2.

01-NOV-2001.

20-APR-2001; 2001W0-US12865.

21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

(CORI-) CORIXA CORP.

Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

L'etisonneuve J, Zhang Y, Jen S, Carter D;

WPI: 2001-616774/71.

N-PSDB: AAS59514.

Example 1; SEQ ID No 2506; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting the sample with a binding agent that binds to the proteins of the invention

CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

xx Sequence 379 AA;

Alignment Scores:

Pred. No.:	1.5e-10	Length:	379
Score:	203.00	Matches:	95
Percent Similarity:	37.89%	Conservative:	63
Best Local Similarity:	22.78%	Mismatches:	155
Query Match:	7.35%	Indels:	104
DB:	23	Gaps:	17

US-10-014-896-1 (1-1509) x ABBA47976 (1-379)

QY 226 CTGGCTGAGTTCGGAATAATCATTAAGTCTTCTTACAGTGTGACACACAGCTTT 285
 DB ||||| |||
 DB 34 LeuAlaGlnIstgLy-----IleGlnSerGlnLys 43
 QY 286 ATCCAGCATGAGTCGTGGAAGATATGACCACTGTCATATCCAGGCTGGACCCC 345
 DB ||||| |||
 DB 44 ValGlnIstgPhe---LeuAspArgAlaSerLeuValSerGlnIleGlnSerGln 62
 QY 346 AGCTTTCAGGCTTACCTGTATGAGTGCATCTTGTATGAGTGCCTGCTGAA---GAA 402
 DB :::: |||
 DB 63 LysValLeuAlaPhe-----SerGlnIstgMetAspArgValAlaSerValSer 80
 QY 403 GGCCTGAGAGTGGCCCATCTCTGGGTTGAGCGGATGAGCGCTCATATGCTGGGGC 462
 DB ||||| |||
 DB 81 LysTrpLysPheProPheGlnAlaThrGlnIstgLeuLysLeuLysGln 100
 QY 463 ACACTGGACGACAGACATCTGTGATGACATTTACCTGACGCGCTGCTGCTGATC 522
 DB ||| |||||
 DB 101 AlaThrAspMetLysSerGlnLeuAlaIstgMetValIleAlaMet----- 115
 QY 523 AGGAAGTACATCCCGAGATCTTCTTCAATTTCTGTGGCCATGATGAGATCATCA 582
 DB ||| ||| |||||
 DB 116 -----IleGlnLeu---HisGlnIstgGlnLys 124
 QY 583 -----GGGACAGGGGCTCAGAGATCTCAGCCCTGCTACAG 618
 DB ||| |||
 DB 125 LeuAsnGlnLysIleArgLeuLeuAlaThrValGlnIstgLeuGlnLysAla 144
 QY 619 TCAAGGGGCTGACAGCTTCAATTCATTTGATGAGGAGGGGCTTCAATTCGATGATTC 678
 DB :::: |||
 DB 145 GlnGlnLeuThrGlnLysGlnIstgAlaAspAspLeuAspIleLeuGlnLys--- 163
 QY 679 ATTCCTTACTTCAAGAGCCCATGCGCTGATTCAGATCTCAGAGAAGGCTTCATGAAC 738
 DB ||||| |||
 DB 164 ---ProSerGlnIstgPhe-----IleValLysAlaIstgGlnSerIleAsn 178
 QY 739 CTGATGCTGCAATGAATGATGATTCAGGCGCATCTTCAGCTCTCCAAAGAGACAGC 798
 DB :::: |||||
 DB 179 TyrThrValLysSerThrGlnLysAsnAlaHisSerMetPro----- 193
 QY 799 ATTGGCATCTTCAGCTGCTGACCGCATTCGAGACACACCAATGCTATCATATTT 858
 DB ----- 193
 QY 859 GGAAGGGAGGAGTGTGACTGATTCAGCAACTGCAAAATGATTTCCCTCCCTGTC 918
 DB -----GlnPheGlnLysAl 197

QY 919 AATATATCTGTCAGCAACCCATGCTATTTGAACCACTATATAGAGGTTTATGAGAGA 978
 DB ||| ||| |||||
 DB 198 AsnAlaIle---AspAsnLeuLeuLeuPheTyrAsnGlnValGlnLysPheValLysSer 216
 QY 979 ATTCCTTACCAATGCAATTAATCAAGACC---ACCAGGCACTCAGCATATTCAAACA 1035
 DB ||||| |||||
 DB 217 IleAspAlaThrAsnGlnIstgLeuGlnLysAspPheIleHisAsnValThrValIleAspGly 236
 QY 1036 GGGGTCAGTCAATGATGATCCCGGAGGCTGGGCGGACAGCAAGTCACTCCGATTCAC 1095
 DB ||| ||| |||||
 DB 237 GlnAsnGlnValAsnSerIleProGlnLysAlaGlnLeuGlnIstgValIleAsnLysSerIle 256
 QY 1096 CCT-----GACAGACAGTCCAGAGTCTTA---GAATCAGCAAGCAATGTGTG 1143
 DB ||| ||| |||||
 DB 257 ProGlnMetAspAsnGlnThrValLysGlnValIleValLysIleIleAsnLysLeuAsn 276
 QY 1144 GCTGATACAGATCCAGTCCATGTTGATGAGTGCCTTGAACCCCTCCCGTC----- 1197
 DB ||| ||| |||||
 DB 277 LysGlnGlnLysAsnValAsnLeuGlnIstgLeuPheAspTyrAspLysGlnProValPheSer 296
 QY 1198 ---AGCCCTTCTGATGACAAAGGCTTGGGCTACAGCTGCTCCGACAGCCGATCAATCC 1254
 DB ||||| |||||
 DB 297 AspLysAsnSerAspLeuValHisIleAlaLysSerValAlaSerAspIleValLysGln 316
 QY 1255 GTCTTCCGGAAGTCAATATTAATCTGCCCCAGTTACTCTATTTGCAACAGACAGCCGA 1314
 DB ||| ||| |||||
 DB 317 GlnIleProLeuIstgGlnLysSer-----GlyThrThrAspAlaIle 330
 QY 1315 TTCTTTCAAAACCTCAGCAGCTGCACTTACAGGTTCTTACCCCATTCATACAGCTGAA 1374
 DB ||||| |||||
 DB 331 GlnPheThrLysAlaLys-----LysGlnPheProValIleIlePheGlnPro 346
 QY 1375 GACTTCAACGCGCATGAGTCAACGCAAGAAATCTGAGTCAAGCAAGCTAT 1425
 DB :::: ||| |||||
 DB 347 GlnAsnGlnThrProHisGlnValAsnGlnAsnValSerIleGlnLysThr 363

Search completed: June 27, 2003, 10:57:55
 Job time : 101.5 secs


```

OY 268 GTGTGACGACACCTTATTCACAGATGAGTGTGAGAGATATACCACTGTTCACT 327
   ::::: ||||| :::::
Db 72 -----AlaHisLysAlaMetThrArgLysValValAlaGlyHisLysLeuValTyrThr 89
OY 328 ATCCAGAGCTGGAGACCCAGCTGTGACACCTGCTGATGAGTGTGATGAGTGTGAG 387
   ||||| ::::: ||||| :::::
Db 90 TrrHrGlySerAsnProAlaLeuAlaProIleValLeuMetLysHisLysGlnAspValVal 109
OY 388 -----CCTGCCCTGAGAAAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 438
   ||||| ::::: ||||| :::::
Db 110 ProValThrProGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 129
OY 439 GATGCCCTCATCTATGCTGCGGACACTGTGACAGCAAGATCTGTGATGATGATGAT 498
   ||||| ::::: ||||| :::::
Db 130 AspGlyLysValTrrGlyArgGlyAlaLeuAspLysSerLysLeuValThrIlePhe 149
OY 499 CAGGCTTGTGAGCTCTGCTGATCAGAAATGATCACTCCGAGATCTTCTTCATTTCT 558
   ::::: ||||| ::::: ||||| :::::
Db 150 GluAlaLeuGlnSerValAlaAlaGlyGlyPheLysProValArgThrValIleIleVal 169
OY 559 CTGGGCTCATGATGAGAGATCATCAGGAGGAGGCTGTGAGATCTGACCTGTGATG 618
   ||||| ::::: ||||| :::::
Db 170 SerGlyLysAspGlyGlyLysValArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 189
OY 619 TCAAGGGGCTGACCTGACCTGATGCTGAGCAGAGGAGGCTGTGATGATGATGATG 678
   ||||| ::::: ||||| :::::
Db 190 SerArgAsnIleLysValGlnPheValLeuAspLysGlyGlyGlyGlyGlyGlyGly 209
OY 679 ATTCCTTAATTCAGAAAGCCATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 738
   ||||| ::::: ||||| :::::
Db 210 ---ProValThrAsnGlnProAlaIleIleGlyValAlaGlyLysGlyTyrAlaThr 228
OY 739 CTGATGCTGACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 795
   ||||| ::::: ||||| :::::
Db 229 LeuLysValThrAlaProAlaValGlyLysSerSerAlaProProLysAspGlyGly 248
OY 796 AGCATTTGGCATCTTGTGACGCTGCTGTGACCGCATTTGGACAGACCAATGCTTAT 855
   ::::: ||||| ::::: ||||| :::::
Db 249 GlyValValThrLeuSerLysLysValGlnAlaLeuHisAspAsnProPheProMetLys 268
OY 856 TTTGGAGGCGGAGTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGCT 915
   ||||| ::::: ||||| :::::
Db 269 Phe---GlnGlyProGlyLysAlaAspMetLeuLysAlaIleSerProHisAlaSerProVal 287
OY 916 GTCATATATATCTGAGCAACCATGCTATTTGAACCATATATATATATATATATAT 975
   ||||| ::::: ||||| :::::
Db 288 ValLysValPheAlaIleAsnThrPheLysSerLysLeuValLysValThrAla 307
OY 976 AGAATCCCTTAACCAATGATATATGAGCAACCAACCAACCAACCAACCAACCAACCA 1035
   ::::: ||||| ::::: ||||| :::::
Db 308 LysSerProAlaGlyAlaAlaMetLeuHisThrThrIleAlaProIleMetLeuLysGly 327
OY 1036 GGGGTCAAGTTCATGTCATCCCGCCAGTGGCCAGGACAGTCAATCTCCGATTCAC 1095
   ||||| ::::: ||||| :::::
Db 328 SerProLysGlnAsnValLeuProGlnAspAlaThrAlaThrIleAsnThrArgIleAla 347
OY 1096 CTTGAGACAGACAGTCCAGAGGCTCTTAAGTCTACAGAAAGCATTTGGTGGTGA 1155
   ||||| ::::: ||||| :::::
Db 348 ProGlyLysPheSerSerAspLysValMetAlaLysValLysGlnValGlyAspLeuPro 367
OY 1156 GTCCAGTTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1200
   ||||| ::::: ||||| :::::
Db 368 ValGlnLeu-----AlaPheGlnGlyHisAspGlnGlnProSerAlaValSer 383
OY 1201 CCTTGTGATGAC-----AAGGCTTGGGCTACACAGTCTCCGCGACAGCCGTACAG 1251
   ::::: ||||| :::::
Db 384 SerThrThrSerAspAlaThrLysThrLeuAlaGlyLeuAlaAlaAspGlnSerGlnAla 403
OY 1252 TCCGCTTCCGGAGATGATATATGCTGCGCCAGTATCTTATTTGAGCAACAGAGACG 1311
   ||||| ::::: ||||| :::::
Db 404 ProValValProGlyLeu-----ValThrAla-----GlyThrAspSer 416
OY 1312 CGATTCCTTTACAACCTCACACAGCTGACATGACAGTTCTACCCCATCTACATACAGCT 1371

```

```

Db 417 ArgTyrMetGlyGlyValLysSerSerAspValTyrArgPheGlnProLeuValLeuThrVal 436
OY 1372 GAGAGCTTCAACAGCATTCATGAGTGCATGACAGAGAAATATCTCACTCAACCTATGAGACC 1431
   ::::: ||||| ::::: ||||| :::::
Db 437 AspGlyThrLysValIleHisLysGlyThrAspGlnHisLysSerLeuAspAsnValGlnLys 456
OY 1432 CAATGAAATTCATCTTGAAGTTCATGATGATGATGATGATGATGATGATGATGAT 1467
   ||||| ::::: ||||| :::::
Db 457 MetValArgPheTyrGlnArgLeuValGlnThrAla 468

```

RESUL 2
 S16693
 Gly-X carboxypeptidase (RC 3.4.17.4) precursor - yeast (*Saccharomyces cerevisiae*)
 MAltName names: carboxypeptidase yscs; protein J0510; protein YUL172w
 CSpecies: *Saccharomyces cerevisiae*
 CDate: 28-May-1993 #sequence_revision 28-May-1993 #text_change 29-Oct-1999
 CAccession: S16693; S16881; S56955; S53878
 R.Bordallo, J.; Bordallo, C.; Gascon, S.; Suarez-Rendules, P.
 FEBS Lett. 283, 27-32, 1991
 A>Title: Molecular cloning and sequencing of genomic DNA encoding yeast vacuolar card
 A.Reference number: S16693; MUID:91243870; PMID:1709881
 A.Accession: S16693
 A:Molecule type: DNA
 A:Residues: 1-576 <BOX>
 A:Cross-references: EMBL:X63068; NID:93593; PIDN:CAA44790.1; PID:93596
 R:Spotman, D.O.; Helm, J.; Wolf, D.H.
 Eur. J. Biochem. 197, 399-405, 1991
 A>Title: Carboxypeptidase yscs: gene structure and function of the vacuolar enzyme.
 A.Reference number: S16881; MUID:91241332; PMID:2026161
 A.Accession: S16881
 A:Molecule type: DNA
 A:Residues: 1-386, 'T', 388-576 <SPO>
 A:Cross-references: EMBL:X57316; NID:93593; PIDN:CAA40571.1; PID:93594
 R:Berthaler, B.; Piravandi, E.; Rinke, M.; Domdey, H.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S56937
 A.Accession: S56935
 A:Molecule type: DNA
 A:Residues: 1-576 <BOX>
 A:Cross-references: EMBL:Z49447; NID:91008366; PIDN:CAA89467.1; PID:91008367; MIPS:YJ
 R.Bordallo, J.; Suarez-Rendules, P.
 Mol. Gen. Genet. 246, 580-589, 1995
 A>Title: C1s and trans-acting regulatory elements required for regulation of the CPS1
 A:Reference number: S53878; MUID:95214618; PMID:7700231
 A.Accession: S53878
 A:Molecule type: DNA
 A:Residues: 1-78 <BOX>
 C:Genetics:
 A:Gene: SGD:CP51
 A:Cross-references: SGD:S0003708; MIPS:YUL172w
 A:Map position: 10L
 C:Function:
 A:Description: hydrolase; metallo-carboxypeptidase
 C:Keywords: glycoprotein; hydrolase; metallo-carboxypeptidase; yeast vacuole
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-56/Product: Gly-X carboxypeptidase #status predicted <MAT>
 F:88,176,228,381,525/Binding site: carboxydate (asn) (covalent) #status predicted

Alignment Scores:
 Pred. No.: 1,896-33 Length: 576
 Score: 540.00 Matches: 159
 Percent similarity: 47.49% Conservative: 78
 Best Local Similarity: 31.86% Mismatches: 184
 Query Match: 19.55% Indels: 78
 DB: 2 Gaps: 19

US-10-014-896-1 (1-1509) x S16693 (1-576)

```

OY 157 GAGGCGCGAAGAGCGCATCCAGATTCACAGTGAAGTTCATGCTGAGAGTCCAT 216
   ||||| ::::: ||||| :::::
Db 92 GlnLysLeuSerAsnAlaValArgIleProThrValVal-----GlnAspLysAsnPro 109

```

```

QY 217 ACTAGAGCC-----CTGGCTGATGTTGGAAAAATCATTCAT 252
Db 110 AsnProAlaAspAspProAspPheTyrGlyHisPheTyrGluLeuHisAspTyrPheGlu 129
QY 253 AAAGTCTTTCTCAGAGGAGGTCACACACAGCTTTATCCAGATCAAGTCCGGAAAGGTAT 312
Db 130 LysThrPheProAsnIle-----HisLysHisLeuLysLeuGluLysValAsnGluLeu 147
QY 313 AGGCACTGTTTCACTTCCAGAGGCTGGACCCAGGCTTACGCTTACGCTGATGGCT 372
Db 148 GlyLeuLeuTyrThrTyrGluGlySerAspProAspLeuLysProLeuLeuMetAla 167
QY 373 CACTTGATGTGGTGGCTGCCCTGAGAA-----GGCTGGAGGTGCCCATTC 423
Db 168 HisGlnAspValValProValAsnAsnGluThrLeuSerSerTyrLysPheProPhe 187
QY 424 TCTGGG-----TTGAGCGGTGATGGGCTCATGATGCTGGGGGACACAGTACGAC 474
Db 188 SerGlyHisTyrAspProGluThrAsp---PheValTyrGlySerAsnAspCys 206
QY 475 AAGAACTGTGTGATGCACTTACTGACGCTTGGAGCTTGGATGATGACAGATACATC 534
Db 207 LysAsnLeuLeuIleAlaGluPheGluAlaIleGluLeuLeuIleAspGlyPheLys 226
QY 535 CCCCAGAGATCTTCTCATTTCTGTGGCCATGATGAGAGATCATCAGGACA--GGG 591
Db 227 ProAsnTyrThrIleValMetSerLeuGlyPheAspGluGluAlaSerGlyThrLeuGly 246
QY 592 GCTCAGAGATTCACAGCCCTGCTACATCAAG-----GGCGTCCAGCTAGCC 639
Db 247 AlaAlaSerLeuAlaSerPheLeuHisGluTyrGlyLysAspArgLysIleTyr----- 264
QY 640 TTCATGTGACGAGGGGGCTTCACTTG-----GATGATTCATCTCTAAC 687
Db 265 SerIleLeuAspGluGlyGluGlyIleMetGluValAspLysAspValPheValAlaThr 284
QY 688 TTCAAGAGCCCATCGCTGATGTCAGTCTCAGAGAGAGGTTTCATGAACCTCATCTG 747
Db 285 -----ProIleAsn-----AlaGluLysGlyTyrValAspPheGluVal 297
QY 748 CAAGTAAACATGACTTCAGGCCACTTTCAGCTCTCCAAAGAGACAGATGCGCATC 807
Db 298 SerIleLeuGlyHisGlyGlyHisSerSerValProProAsnIleThrIleGlyIle 317
QY 808 CTTCAGCTGCTGTACGCCGATTTGAGCAGACACACCAATGCCATCATTTTGAAGCGG 867
Db 318 AlaSerGluLeuIleThrGluPheGluAlaAsnProPheAspTyrGluPheGluPheAsp 337
QY 868 ACACTGGTGACTGATTTGACGACAACTGCGAAATAGTTCCCTGCTCATATATATC 927
Db 338 AsnProIleTyrGluLeuLeuThrCysAlaAlaGluHisSerLysSerLeuSerLysAsp 357
QY 928 CTGAGCAACCCATGGCTATTGACCACTTATAGCAGG----- 966
Db 358 ValLysLysThrIleLeuGlyAlaProPheCysProArgLysAspLysLeuValGlu 377
QY 967 TTTATGGAGAGAAATCCCTTAACCAATGCAATTAATCAGAGACACCGCACCTACCATTA 1026
Db 378 TyrIleSerAsnGlnSerHisLeuArgSerLeuIleArgThrGlnAlaValAspIle 397
QY 1027 TTCAAAGCAGGGGTCAATGATTCATCCGCCACAGGGGCCACGCGCAAGTCAATTC 1086
Db 398 IleAsnGlyGlyValLysAlaAsnAlaLeuProGluThrThrArgPheLeuIleAsnHis 417
QY 1087 CGGATTCACCTTGACAGACAGCTCCAGAGCTCTAGAA----- 1125
Db 418 ArgIleAsnLeuHisSerSerValAlaGluValPheGluArgAsnIleGluTyrAlaLys 437
QY 1126 -----CTCAGAGAGAG-----ATTGTGGCT 1146
Db 438 LysIleAlaGluLysTyrGlyTyrGlyLeuSerLysAsnGlyAspAspTyrIleIlePro 457
QY 1147 GAT-----ACAGAGTCCAGTTCATGTGTTGATGCTTTCAGACCCCTCCGCTC 1197

```

```

Db 458 GluThrGluLeuGlyHisIleAspIleThrLeuLeuArgGluLeuGluProLeu 477
QY 1198 AGCCCTTTCATGAGACAGGCTTGGGCTTCAGATGCTGCCGACAGCTACAGTCCGTC 1257
Db 478 SerProSerSerGlyProVal-----TrpAspIleLeuAlaGlyThrIleGlnAspVal 495
QY 1258 TTC-----CCGAAAGTCAATATTACTGCCCATCTACTCT 1293
Db 496 PheGluAsnGlyValLeuGlnAsnAsnGluGluPheTyrValThrThrGlyLeuPheSer 515
QY 1294 ATTGGCAACAGACAGCCGATTTTCACAACTCACCACATCGCATCTACAGTTCTAC 1353
Db 516 ---GlyAsnThrAspThrLysTyrTyrTrpAsnLeuSerLysAsnIleTyrThrPheVal 534
QY 1354 CCATCTACATACAGCTCAGACATCAACATCAGATCAGTACAGAGAAATCTCA 1413
Db 535 GlySerIleIleAspIleAspLeuLeuLysThrLeuHisSerValAsnGluHisValAsp 554
QY 1414 GTCCAAAGCTTATGAGACCCAGTGAATTCATCTTGAATGATTCAGAAATCTGAC 1470
Db 555 ValProGlyHisLeuSerAlaIleAlaPheValTyrGluTyrIleValAsnValAsn 573

RESULT 3
T38349
carboxypeptidase s precursor - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_rev199 #text_change 03-Dec-1999
C/Accession: T38349
R/Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
Submitted to the EMBL Data Library, August 1995
A/Reference number: Z2187
A/Accession: T38349
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-596 <DJB>
A/Cross-references: EMBL:298601; PIDW:CAB11265.1; GSPDB:GN00066; SPDB:SPAC24C9.08
A/Experimental source: strain 972h-; cosmid c24C9
C/Genetics:
A/Gene: SPDB:SPAC24C9.08
A/Map position: 1

Alignment Scores:
Pred. No.: 8,65e-33 Length: 596
Score: 531.50 Matches: 149
Percent Similarity: 40.75% Conservative: 91
Best Local Similarity: 25.30% Mismatches: 212
Query Match: 19.24% Indels: 137
Gaps: 13

US-10-014-896-1 (1-1509) x T38349 (1-596)
QY 28 GCCCTGTGCGTATGCTGCTCTAGTTTCCCTACCGCTCCAGATCGATGGCCCGAG 87
Db 40 AlaPheValGlyPheLeuIleLeuIlePheValTyrThrLeuLeuArgGlyGlySerAsn 59
QY 88 AGCGGGAGCATCAAAAGGCGGTGCGGAATCCCTTCAGTTTCGCAAGAGAGAGAGCGGTC 147
Db 60 AspAsnAspLysGlnGluMetSerHisSerProGlySerCysMetAspSerGluSerAla 79
QY 148 GCGATGAAAAGGCGGTGAAA----- 168
Db 80 AlaValSerThrSerAlaLysCysTyrIleProProValLeuThrProAlaLysGluPro 99
QY 168 ----- 168
Db 100 LysLeuGlyAspAspValSerGlyIleAspTyrIleArgSerProGluPhePheAsnAsp 119
QY 169 -----GGTCCATCCAGATTCACAGACTGCTTTTAC----- 201
Db 120 SerLeuValArgPheGlnGluLeuLeuArgIleProThrValCysTyrTrpAspMetGly 139
QY 202 -----TCTGGAAGTCCAAATACTACAGCCCTGCTGATGTTTCGAAATATCATTTATAA 255

```

```

Db      140 AspValGlyAspAspArgPheAspIlePheIleValIlePheGlnAspIleValArgGlu 159
      256 GTCTTCTACAGAGGTCAGACACAGCTTATCCAGATAGAGTGGAGAGTATAGC 315
      160 LeuYrProAsnIlePheIleValS-----LeuIleValGluYrValAsnThrTyGly 177
      316 CACGTGTCATATCAGAGGTCAGACCCAGCTGAGCCCTGACCCGTCATAGGTCAC 375
      178 LeuLeuIleThrLeuGluGlySerAsnIleAspLeuIleValIleMetGlyHis 197
      376 TTGATGTGTGCT-----GCCCTGAGAGAGCTGAGAGTGGCCCATTTCT 426
      198 GlnAspValValProValAsnGlnIleSerLeuAspIleGlyTrpIleProIleSer 217
      427 GGGTTGAGCGCTGATGGCGTCATATAGTCGGGGCAGCTGAGCAGAGAACTGTG 486
      218 AlaThrTyHisAsnGlyHisValTySerArgIleValIleAspAspIleValSerVal 237
      487 ATGCGATTAATCAGAGGCTGAGAGCTGCTGTCATGAGAGTACATCCCGAAGATCT 546
      238 ValAlaIleLeuGlnIleValLeuGlnIleValIleSerAspTyTrpIleProGluGlnThr 257
      547 TTCTTCATTTCTCTGGCCATGATGAGAGTATCAGG--ACAGGGCTCAGAGGATC 603
      258 ValIleAlaSerPheGlyPheAspGluGluValSerGlyTyTrpGlyIleLeuProIle 277
      604 TCAGCCCTGCTACAGTACAG-----GGGCTCAGAGCTGAGCTTATGTTGGAC 651
      278 AlaIleIleValLeuYrGluArgTyGlyLysAspIleVal-----AlaLeuIleLeuAsp 295
      652 GAGGGGGCTTACATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
      296 GlnGlyGlyPheThrIleAsnLeuPheGlyThrLeuPhe-----AlaThrVal 311
      712 GCACTCAGAGAGAGGTTCCATGAACTGATGCTGCAACTGAACTGATGAGGCCAC 771
      312 CysValAlaGlnIleValGlyTyTrpMetAspValIleIleValLeuYrThrProGluGlnHis 331
      772 TCTTCAGCTCTCCCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 831
      332 AlaSerIleProProIleAsnIleValIleGlyLeuMetSerIleValIleThrGlnIle 351
      832 GAGCAGACCAATGCTATCATATTTGAGAGCGGAGACGTGCTGATGATGAGCAAA 891
      352 GlnGlu----- 353
      892 CTGGCAAAATGATTTCCCTTCCCTGTCATATATATCTTGAACACCCATGGCTATTGAA 951
      354 -----ProPheGlyGlyIleuThrPheGluAsnPro-----PheTy 366
      952 CCACTTATAGAGAGTTTATGAG----- 975
      367 ThrThrLeuGlnCysPheAlaGluAsnSerAlaAspMetAspAsnLeuArgGlnLeu 386
      976 -----AGAAATCCCTTAACCAAT 993
      387 IleIleSerGlyAspThrGluMetIleAspLeuPheSerIleSerIleArgLeuTyArg 406
      994 GCATTAATCAGAGCCACAGGCTACCATATTTCAAAGAGGGCTCAAGTTCAATGTC 1053
      407 TyrPhePheGluThrSerIleAlaValAspValIleAsnGlyIleValLysValAsnAla 426
      1054 ATGCCCCCAGTGGCCAGGCAACAGTCATCCGATTCACCCCTGGAGACAGACAGTCCA 1113
      427 LeuProGluGluThrThrLeuAlaValAsnHisArgValAspAlaSerIleGlyLeuLys 446
      1114 GAGGTCTAGAACTCAGAGAAACATTTGGCTGATACAGAGTCCAGTTCCATGTC--- 1170
      447 GlnValTyAspArgTyGlyGlyLeuLeuGluGluPheGlyHisIleGluThrHisValAsn 466
      1170 ----- 1170

```

```

Db      467 ValThrLeuPheAsnGlyIleThrValValGluTyGluAspAlaIleGlyHisIlePhe 486
      1171 -----TTGAGTCCCTTTGACCCCTCCCGTCAGCCCTTGATGACAGAGCCTTG 1221
      487 AlaSerThrAlaIleYrThrLeuGluProSerProValSerProTyAspIleSerAsp 506
      1222 GGTACACACTGTCTCCCGACAGACCGTACAGTCCGCTTCCCGAAGCAATATTACTGCC 1281
      507 AlaTyLeuLysLeuAlaGlyAlaIleArgTyThrPheGlyAspGlyThrSerValThr 526
      1282 CCAGTACTTCTATTGGCAGACAGACCGCATCTTTACAAACCTGACACAGTGCATC 1341
      527 ProAlaLeuMetProAlaAsnThrAspThrAsnIleGlyTyTrpAsnLeuThrSerAsnIle 546
      1342 TACAGTTCTACCCCATCTACATACAGCTGAGAGCTTCAACAGCATCATGAG----- 1395
      547 TyrArgTyTrpThrProValSerThrAsnSerThrSerLysAsnSerPheAsnGlyHisThr 566
      1396 GTCAACGAGAAATCTCAGTCCAGACCTATGAGACCCAGTGAAATTCATCTTTGAGTTG 1455
      567 IleAsnGluAsnMetArgTyTrpAspAlaHisMetAspSerIleGluPhePheTyAsnPhe 586
      587 IleLeuValSerAspSerGlyGlu 595

```

RESULT 4
T35974
probable aminoacylase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T35974
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21551
A:Accession: T35974
A:status: preliminary; translated from GB/EMBL/DBJ

A:molecule type: DNA
A:residues: 1-443 <SEED>
A:cross-references: EMBL:AL035161; PIDN:CAA22735.1; GSPDB:GN00070; SCOPDB:SC9C7.23
A:Experimental source: strain A3(2)
C:genetics:
A:gene: SCOPDB:SC9C7.23
C:superfamily: succinyl-diaminopimelate desuccinylase

Alignment Scores:

Pred. No.:	9,87e-20	Length:	443
Score:	361.50	Matches:	134
Percent Similarity:	45.63%	Conservative:	75
Best Local Similarity:	29.26%	Mismatches:	181
Query Match:	13.09%	Indels:	68
DB:	2	Gaps:	20

US-10-014-896-1 (1-1509) x T35974 (1-443)

```

      156 AGAGCGCTGAAAGTGGCATCCAGATTTCCAGAGTGAATTTAGCTGTGAGAGTCCA 215
      34 ArgGlyGlyIleAspCysGln-GluArgProAlaAlaGluIleValAlaIleAlaArg----- 51
      216 TACTACAGCCCTGGGTGATGCGGAAATATCATTCATTAAGTCTTTCTCCACAGTGCAG 275
      52 -----LeuAlaGluAlaGly-----IleGlnProThrLeuGlu 63
      276 CACCAAGCTTTATCCAGATGAGATGCTGGAAGATATAGCACCTGTCTACATCCAAAG 335
      63 uArgThrAlaGlyArgThrAsnValAlaIleArg-----IleGlu 77
      336 CTGGAGCCCAAGCTTGAAGCCCTACCTGCTGATGCTGCTGATGCTGCTGCTGCTGCTGCTG 395
      77 TyrAspProSerAlaAspAlaLeuValHisGlyHisIleAspValValProAlaGlu 97
      396 TGAAGAGCTGGAGAGTGGCCCATTTCTGGGTTGGAGCTGATGAGCTCATTTAGG 455

```


Db 219 GluValGlnAlaGlyGlyValTyrGlnAspPheThrValThrAsnProGly 238
QY 766 GGCACCTCTTCAGCTCTCCAAAGAGACAGCATGGCATCTTGACGTGCTCAGC 825
Db 239 GlyHisSerSerArgProValSerProAsnAlaIleTyrGlnLeuSerAlaIleLeuAsp 258
QY 826 CGATTGGAGACAGACCAATGCATTCATTGGAAAGCGGACAGTGGGACTGTATTG 885
Db 259 ArgIleGlyAlaTyrGlnPheProProAlaGlyPheAsnAspAlaThrArgGlyTyrPheThr 278
QY 886 CAGCAACTGGCAATGATGCC-----TTCCTGTCAATTAATTCATGACCAAC 936
Db 279 GlnMetGlnAlaArgValThrProGlnGlnAlaAlaIleMetLysThrLeuValAlaAsp 298
QY 937 CCAATGGTATTGAACCATTAATAGCAGGTATATGAGAGAAATCCCTTAACCAATGA 966
Db 299 -----ValAsnAspProAlaAlaLeuAlaLeuIleThrLysAspArgThrTrpAsnSer 316
QY 997 ATATATCGACACGACGACGACATCCATATTCAAAGCGGGTCAAGTTCAATGTCAATC 1056
Db 317 MetLeuArgThrThrCysValAlaAlaThrMetValSerAlaGlyHisAlaProAsnAlaLeu 336
QY 1057 CCCCAGTGGCCAGGACGACAGTCAATTCGCGATTACCCCTGACAGACAGTCCAAAG 1116
Db 337 ProGlnArgAlaThrAlaAsnIleAsnGlyArgIleLeuProGlyThrProIleAspGlu 356
QY 1117 GTCTTGAACTCAGAAACATTTGGCTGTATACAGAGTCCAGTTCATGTGTAGT 1176
Db 357 ValLysAlaLysLeuThrGlnLeuAlaAlaAspProAlaValAla-----ValThrLeu 374
QY 1177 GCCTTGACCCCTCCCGTCAGCTTCGTGATGACAAAGCCCTTGCGGTACCAAGCTG---- 1233
Db 375 AlnHisSerSerLysProAlaSerProArgProArgProLeuThrProAlaIleMetAlaPro 394
QY 1234 CTCGCCAGACGCTAGCTGCTGCTCCGGAAGTCAATTAATCTGCCCCAGTACTCTT 1293
Db 395 IleGlnLysAsnAlaAlaLysLeuThrProGlyValProIle-----LeuProValMetSer 413
QY 1294 ATTTGAGACAGACAGCCGATCTTTACAAACCTGACACAGTGGCATCTACAGTGTCTAC 1353
Db 414 ThrGlyAlaThrAspAlaValAlaHisThrSerAlaAlaGlyIleProThrTyrGlyValThr 433
QY 1354 CCCATGTACATACAGCTGTAAGACTTCAACAGCATCCAGTGGAGTCAAGAGAAATCTCA 1413
Db 434 GlyLeuPheHisGlyProGluGly-----ThrGlyAlaHisGlyLeuHisMetLysMetLys 452
QY 1414 GTCCAAAGCTATGAGACCCAAAGTGAATTCATCTTGAGTTCATTCAGAT 1464
Db 453 ValLysSerLeuTyrGlnGlyAlaArgAspPheLeuHisGlyLeuIleGlnAsp 469

RESULT 6
B87070
Probable peptidase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 02-Aug-2002
C:Accession: B87070
R:Coile, S.T.; Eiglmier, K.; Parhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Dutfoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; PMID:21128732; PMID:11234002
A:Accession: B87070
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <STO>
A:Cross-References: GB:AL450380; NID:g13093213; PIDN:CA631669.1; GSPDB:GN00147
C:Genetics:
A:Gene: MLI288
C:Superfamily: succinyl-diaminopimelate desuccinylase

Alignment Scores:

Pred. No.: 2,45e-17 Length: 467
Score: 330.50 Matches: 114
Percent Similarity: 46.21% Conservative: 69
Best Local Similarity: 28.79% Mismatches: 166
Query Match: 11.97% Indels: 47
DB: 2 Gaps: 17

US-10-014-896-1 (1-1509) x B87070 (1-467)

QY 328 ATCCAGGCTGCGACCCCGCTTGACGCTTACGCTGATGGCTTGTGATGTG 387
Db 93 LeuAlaGlyGlnAspSerSerThrGlyAlaLeuLeuIleHisGlyHisLeuAspValAl 112
QY 388 CCTGCCCTGAAGAAGCTGGAGGTGCCCTTCTGGGTGGTGAAGCGGCTC 447
Db 113 ProAlaGlnThrAlaGlnThrSerValHisProPheSerGlyAlaValGluGlyGln 132
QY 448 ATCTATGCTGGGCGACACTGACAGACAACTCTGTATGATGCAATTAAGTCAAGCTTG 507
Db 133 ValTrpGlyArgGlyAlaIleAspMetLysAspMetValGlyMetMetIleValAla 152
QY 508 GAGCTCGCTGATGAGGAATGACATCCCGAAGATCTTCTCATTTCTCTGGGCAT 567
Db 153 ArgGlnLeuLysGlnAlaGlyIleAlaProProArgAspLeuValPheAlaPheValAla 172
QY 568 GATGAGAGATCATCAGGACA---GGGCTCAGAGATCTCAGCCCTGTACAGTCAAG 624
Db 173 AspGlnGlnHisGlyGlySerTyrGlySerLys-----TrpLeuValAspAsnArg 189
QY 625 -----GGGCTCAGCTGACCTTCATGTGGACGAGGGGGCTTCATCTTG 669
Db 190 ProAspLeuPheAspGlyValThrGluAla---IleGlyLysValGlyGlyPheSerLeu 208
QY 670 GATGATTCATATCC-----AACTCAAGAAGCCCATCGCTGATGAGTGTCA 720
Db 209 Thr-----ValProCysArgAsnGlyGlyGlnArgGlyLeuIleGluThrAla 226
QY 721 GAGAAAGTTCCATGAACTCATGCTGCGAATTAATGATGATGAGGCACTTTCAGT 780
Db 227 GlnLysGlyMetGlnTrpMetArgLeuThrAlaArgGlyAlaGlnHisGlySerMet 246
QY 781 COTCAAAGAGACAAAGATGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 247 ValHisAsnGlnAsnAlaValThrAlaValAlaGluAlaValAlaAlaAspGlyArgHis 266
QY 841 CCAATGCTCATATTTGGAAGCGGACAGTGTGATGATGATGATGATGATGATGATGAT 900
Db 267 GlnPheProLeuVal-----ThrThrAspThrValAlaGlnPheLeuAlaAlaIleSerGlu 285
QY 901 GAGTTCCCTTCCCTGCTCATATATATATCCCTGAGCAACCCATGCTATTTGAACCACTTATA 960
Db 286 GlnThrGlyLeuGlnPheAspIle-----GlySerProAspLeu-----GluGlyAlaIle 302
QY 961 AGCAGGTTTATGAGAGAAATCCCTTAACCAATGCAATTAATCAGAGCCACACCGGCATC 1020
Db 303 GlnLysLeuGlyProMetAlaArgMetLeuLysAlaValLeuLysTyrAspThrAlaAsnPro 322
QY 1021 AACATATTTCAAGAGAGGCTCAAGTCAATGTCATCCCGACAGGCGGCGACAGTCC 1080
Db 323 ThrValLeuLysAlaGlyTyrLysValAsnValAlaProAlaThrAlaGlnAlaMetVal 342
QY 1081 AACTTCGGGATTCACCTGACAG-----ACAGTCAAGAGTCTTA 1122
Db 343 AspCysArgIleLeuProGlyArgGlnAlaAlaPheGlnAlaAlaIleAspGluLeuIle 362
QY 1123 -----GACATCAGAGAGAACATTTGCTGCTGATTAACAGATCCAGTTCATGTGTGAGT 1176
Db 363 GlyProAspValThrArgGlnThrPheLysAsp----- 373
QY 1177 GCCTTGACCCCTCCCGTCAGCTTCGTGATGACAAAGCCCTTGCGGTACCAAGTGTCTC 1236
Db 374 ---LeuProProTyrGlnThrAlaPheAspGlyAspLeuVal-----AspAlaMet 389

QY 1309 AGCGATTCCTTACAAACCTCCACGATGCTACAGTTCACCCCATCTACATACAG 1368
|||||
Db 407 SerArgTyrPheThr-----ProLeuGlyValIlys 416
QY 1369 CCTAGAACCTTC-----AAACGATCCATGAGTCAACGAAATCTCAGTC 1416
|||||
Db 417 AlAlleAspIleGlyProArgGlyGlyAsnIleHisGlyProAsnGluTyrValGluVal 436
QY 1417 CAAGCC 1422
|||
Db 437 Asper 438

RESULT 13

C96507

hypothetical protein T12C22.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C96507

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huitzer, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C96507

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-438 <STO>

A:Cross-references: GB:AE005173; NID:g8655992; PIDN:AAF78265.1; GSPDB:GN00141

C:Genetics:

A:Gene: T12C22.9

A:Map position: 1

Alignment Scores:

Pred. No.:	3,24e-08	Length:	438
Score:	212.00	Matches:	100
Percent Similarity:	37.04%	Conservative:	80
Best local Similarity:	20.58%	Mismatches:	162
Query Match:	7.68%	Indels:	144
		Gaps:	19

US-10-014-896-1 (1-1509) x C96507 (1-438)

QY 187 ACAGTACTTTAGCTCTGAGAGTCCAAATACT-----ACAGCCCTGGCTGACTTC 237
|||||
Db 14 SerIleIlePheSerIleuGlnSerSerSerGluGlnAspThrProIleThrArgPhe 33
QY 238 GGAATAATCAAT-----249
|||||
Db 34 GlnGlnTyrIleuArgPheAsnThrAlaHisProAsnProAsnTyrThrIleProIleSer 53
250 -----CATTAAGTCTTCTTACAGTGGTGCAGCAGCAGCAGCAGCAGCAGCAG 291
|||||
Db 54 PheIleuIleAsnGlnIleAsnIleGlnIleGlyLeuThrIleThrIleGlnIlePheIleSer 73
292 CATGAGTGGTGGAGAGTATACGACGCTTACACATCCAGAGGCTCGGAGCCGAGCTTG 351
|||
Db 74 GlyIleuProIle-----IleuIleuIleThrIleuIleuIleuIleuIleuIleu 89
352 CAGCCCTGCTGAGTGGCTCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
|||
Db 90 ProSerIleuPheAsnSerIleuAsnSerValProAlaGlnSerGluIleuIleuIleuIleu 109
412 GTGCCCCCATCTCTGCTGGTGGAGCGT---GATGCGCTCATGTAGTGGGCGGAGCAGCTG 468
|||
Db 110 TyrProIleuPheSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 129

QY 469 GAGCAGAAAGAACTCTGTGATGCATTACTGAGAGCGCTTGGAGCTCTGCTGATCAGAG 528
|||||
Db 130 AspAspLysCysIleGlyValGlnTyrIleuGlnSerIleArgAsnIleuLysSerArgIly 149
QY 529 TACATCCCCGAGAACTCTTCTTCACTTCTCTGGGCGCAATGAGAGAGTCAATCAGG--- 585
|||
Db 150 PheSerProLeuArgThrIleHisIleSerTyrValProGlnGlnIleGlyIlePhe 169
QY 586 ACAGGGGCTCAGAGATCTCAGCCCTGCTACAGTCAAGGGCGCTCCAGCTCACTTCAAT 645
|||
Db 170 AspGlyMetMetIlePheAlaIleSerSerGluPheLysAspLeuAsnIleGlyPheAla 189
QY 646 GTGAGCAGAGG---GACTCATCTGATGATGATGATGATGATGATGATGATGATGATGATG 693
|||
Db 190 MetAspGlnGlyGlnAlaAsnProGly-----AspGluPhe-----201
QY 694 AAGCCCATGCGCTTGAATTCAGCTCCAGAGAGGGTTCATGACATGCTGCGCACTA 753
|||
Db 202 -----ArgValPheTyrAlaAspArgValProTyrPheIleValIleValAla 217
QY 754 AACATGACTTCAGCCGACCTCTCAGCTCTCCAAAGGAGACAGCAGATTCGATCCCTGCA 813
|||
Db 218 GlnGlyIleProGlyIleIleValIlys-----226
QY 814 GCTGCTGTCAGCCGATGAGAGACACCAATGCTATCATATTTGGAAGCGGAGCAGTG 873
|||
Db 227 -----LeuTyrAspAsnSerIleAlaIleTyr-----233
QY 874 GTGACTGTATTCAGCAACGTGCAAAAGATTTCCCTCCCTGCAATATTAATCTGAGC 933
|||
Db 234 GluAsnIleuMetLysSerVal-----240
QY 934 AACCCATGGCTATTTGAACCACTTAAGCAGGTTATGAGAGA-----978
|||
Db 241 -----GluIleuIleSerArgPheArgGluSerGlnPheAspPheVal 254
QY 979 -----AATCCCTTA---ACCAATGCA 996
|||
Db 255 LysAlaGlyLysAlaIleTyrSerGluValIleSerValAsnProValTyrLeuLysAla 274
QY 997 ATATCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1056
|||
Db 275 GlyThrProIleThrIle-----GlyPheValMetAsnMetGln 287
QY 1057 CCCCAGAGGCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1098
|||
Db 288 ProSerGluAlaGlnIleGlyTyrAspIleuArgLeuProIleMetAlaAspProAspVal 307
QY 1099 -----GACAGCAGCTCCAAAGTCTTAAGAACTCAGAGAACTT 1140
|||
Db 308 MetLysLysArgIleAlaGlnGluTyrPheAlaProSerIleArgAsnMetThrTyrSerIle 327
QY 1141 GTGGCTGATTAACAGATCCAGTCCAGTGTGAGTGGCTTGAACCCCTCCCGCTCAGC 1200
|||
Db 328 GlnGlnIleuGlyLysLeuArgAspHisIleu-----GlyArgProIle---MetThr 343
QY 1201 CCTTGTATGACAGAGCCCTTGGCTCAGCAGCTGCTCCGACAGCAGCAGCAGCAGCAGCAGCAG 1260
|||
Db 344 ProValAsnAspSerAsnProTyrPheSerIlePheLysGlnAlaValGlnAlaMetCyl 363
QY 1261 CCGAGAGCAATATATCTGCCCCAGTCACTTCTTAATGGCAACAGCAGCAGCAGCAGCAGCAG 1320
|||
Db 364 GlyLys-----LeuAlaIleuArgProGluIleuAlaIleuIleuIleuIleuIleuIleu 381
QY 1321 ACAAACTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1380
|||
Db 382 ArgThrLeuGlyIleProThrPheIlePheSerProMetThrAsnThrPro-----398
QY 1381 AAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1419
|||
Db 399 IleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 418
QY 1420 GCGTATGAGACCCAGAGTG 1437

Db 419 ValTyrGluserValIle 424

|||||::: :::

RESULT 14

A96506

probable aminoacylase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96506

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96506

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1,435 <STO>

A:Cross-references: GB:AE005173; NID:g11120814; PIDN:AA930993.1; GSPDB:GN00141

A:Gene: T7O23.14

A:Map position: 1

Alignment Scores:

Pred. No.: 5.5e-08 Length: 435

Score: 209.00 Matches: 94

Percent Similarity: 38.96% Conservative: 63

Best local Similarity: 23.33% Mismatches: 161

Query Match: 7.57% Indels: 85

DB: 2 Gaps: 15

US-10-014-896-1 (1-1509) x A96506 (1-435)

QY 319 CTGTTCACATATCAAGGCTCGGACCCAGCTTGCAGCCCTTACCTGTGATGAGCTTCACTTT 378

||| |||||::: ::||| ||| ::|||

Db 81 LeuLeuThrTrpLeuGlySerAlaLeuAlaLeuSerIleLeuPheAlaSerHisIleu 100

379 GATGTGGTGGCTGGCCCTCGAAGAGGCTGGAGGCTGCCCATCTCTCGGGTTGGAGCGT 438

||| ||||| ||||| ||| ||| ||||| ||||| |||||

Db 101 AspSerValProAlaGluSerGlnuyluStrpIleHisProPheSerAlaHisAlaTrp 120

439 --CATGGCTCATCTATGCTGTGGGGGCACATGGACGACGAACATCTGTGATGACATTA 495

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 121 IleAspGlnHisIleTrpAlaArgGlyAlaGlnAspArgLysGlyValGlnTrp 140

496 CTGCAGGCTTGGAGAGCTCTGCTGTGATCAGAGAGATACATCCCCGAGATCTTTTCTTCA 555

||||| ||||| ||| ||| ||| ||| ||| ||| |||

QY 141 LeuGlnAlaIleLeuArgAsnLeuLysSerTrpSerProLeuLysTrpIleHisIle 160

556 TCTCTGGGCACTGATGAGAGTCAATCAGGACAGAGGCT---CAGAGGATCTCAGCCCTG 612

||| ::||| ||| ||| ||| ::|||

Db 161 SerTrpValProGlnGlnuyluIleGlyLysMetMetLysPheAlaIleAsp 180

613 CTACAGTCAGAGGCGCTCCAGCTAGCCCTTCAATGTGTGACGAGGG-----GGC 660

::: ::: ||| ||| ::||| ||||| |||||

Db 181 SerGlnPheLysAspLeuAlaSerLeuGlnPheValMetAspGlnuyluGlnAlaSerPro 200

661 TTCAATCTGGAGAGGATGATCTTCACTTCAAGTTCAGAGGCCATCGGCTGATGACAGTCA 720

||| ||||| ||||| ||||| ||||| ||||| |||||

Db 201 ---AspGlnPhe-----ArgValPheTrpAla 208

721 GAGAGGCTTCCATGACATCTCATGCTGCGACATTAACATGACTAGGCCATCTTCACT 780

||||| ::||| ::||| ||| ||||| :::

Db 209 GlnuyluTrpTrpTrpHisLeuAlaIleArgLysAsnGlnuyluMetProGlnuyluAlaLys 228

781 CTTCCAAAGGACACAGATGGCATCTTTCAGAGCTGTGTC-----AGCGATTTG 831

||||| ::||| ||| |||||

Db	229	Leu1y7rAspsnsrSer1aMeG1uAsnLeuMeLysSerValG1uLeu1IeSerArgPhe	248
Qy	832	GAGAGACACCAATGCGCTATCATATTGTTGGAACCGCGGACAGGGGACTGATATTCACCAA	891
Db	249	Arg1uSerG1nPhesAprheValLysAlaGly-----Lys	260
Qy	892	CTGGCAATGAGTTTCCCTTCCTCGTAATTAATTCCTGAGCAACCCATGGCTATTGAA	951
Db	261	Ala1aIasnsr-----G1uVal1IeSerValAsnPro-----	271
Qy	952	CCACTTATACAGCGTTTATGGAGAGAATTCCTTAACCAATGCAATTAATCAGACCAACC	1011
Db	272	-----Val1yLeu1yLysAlaGly	277
Qy	1012	ACGGCAGCTACCATATTCAAGCAGGGGTCAAGTTCAATGTCATGCCCCCGAGTGGCCAG	1071
Db	278	ThrProSerThr-----ThnG1yPheValMeLysnMetG1nProSerG1uAlaG1u	294
Qy	1072	GCCACAGTCAACTTCGGATTACCCCT-----	1098
Db	295	ValG1yTy7rAspLeuArgLeuProProMeta1aAspProVal1a1aLeuLysArg1Ie	314
Qy	1099	---GGACAGACAGTCCAAAGAGGCTCTGAAGTCAAGCAAGAACATTTGGCTGATACAGA	1155
Db	315	AlaG1uG1nProAlaProSerSer1IeArgAsnMetThyTy7rThrLeuLysG1nG1nG1yLys	334
Qy	1156	GTCCAGTTCATGTGTGTGAGTGCCTTTGACCCCTCCCTCCCGTCAGCCCTTCGATGACAG	1215
Db	335	LeuThy7rAspHisLeu-----G1yArgPro1IeMetThrThrThrAsnAsp-ThrAs	351
Qy	1216	GCGTTGGGCTTACCAAGCTCTCTCCGACAGCCGTACAGTCCGCTTCCCGGAATCAATTT	1275
Db	351	nPrOT7r---TrpSer1IePheLeuG1aG1n1aValaG1uAlaThrG1yG1yLys-----Leu	368
Qy	1276	ACTGCGCCAGTACTCTTCAATTGGCAACACACAGACCGCATCTTTACAAAGCTCAGCACT	1335
Db	369	Ala1yAspProG1u1IeLeu1IeSerThrThrAspSerArgPhe1IeArgThrLeuG1y1Ie	388
Qy	1336	GGCATCTACAGGTTCTATCCCATCTACATACAGCCTGANAAGCTTAACAGCATCCAGGA	1395
Db	389	ProVal1eug1yPheSerProMet1IeAsnThrPro-----1IeLeuValHisAsp	405
Qy	1396	GTCACAGCAAAATC-----TCAGTCCAAAGCTATGAGACCCAA	1434
Db	406	HisAsnG1uPhelLeuLysAspThyValPheMetLysG1y1IeG1yValTy7rG1uSerVal	425
Qy	1435	GTG 1437	
Db	426	Ile 426	
RESULT 15			
T19180			
Hypothetical protein C10C5.3 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
C:Accession: T19180			
R:Matthews, P.			
submitted to the EMBL Data Library, December 1995			
A:Reference number: Z19085			
A:Accession: T19180			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-399 <MIL>			
A:Cross-references: EMBL:Z68214; PIDN:CA92445.1; GSPDB:GN00022; CESP:C10C5.3			
A:Experimental source: clone C10C5			
C:Genetics:			
A:Gene: CESP:C10C5.3			
A:Map position: 4			
A:introns: 27/1; 48/3; 215/3; 268/1; 303/3; 330/2; 350/3; 383/1			
Alignment Scores:			
6,99e-08 Length: 399			
Pred. No.: 207,50 Matches: 109			

Percent Similarity: 38.65% Conservative: 68
 Best Local Similarity: 23.80% Mismatches: 188
 Query Match: 7.51% Indels: 93
 DB: 2 Gaps: 18

US-10-014-896-1 (1-1509) x r19180 (1-399)

```

OY 136 GAGAAAGCGTCGCGATG-----AAGAGGGCGTGAAGGTCGATCCAGATTCCA 186
    ||||| :||||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 3 GIUGLUNHSLIEALVALTHrArpRheArpGluTyrLeuArpValAsnThrGIUGLrPro 22
OY 187 ACAGTACTTTAGCTGTGAGAGTCCAAATCTACAGCCCTGGCTGAGTGGGAAATAC 246
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 23 AsnProArpTyrAlaAlaCysArpAspRheLeuRheTyrTyrAlaAspGluLeuGlyIle 42
OY 247 ATTCATAAGTCTTTCCTACAGTGTGACACCCAGCTTTATCCAGCATGAAGTCGTGAA 306
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 43 AlaArpArpSerPheGluThrValProGluAlaIlePheVal----- 56
OY 307 GAGTATAGCCACCTGTCATATCCAAAGGCTCGAGCCGAGCCGATTCAGCCCTACTGCTG 366
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 57 -----IleMetThrIleProGlySerGlnProGluLeuProSerIleMetLeu 72
OY 367 ATGGCTCACTTTGATGTGTGCTGCTGCCCTGAGAGAGGCTGGAGAGTGGCCCATTTCT 426
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 73 TyrSerHisThrAspValValProThrPheArpGluHisThrIleAspProTyrSer 92
OY 427 GGGTTG---GAGCTGATGGCGGTATCTATGTGTGGGCGACATCGACAGACAAGACTCT 483
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 93 AlaPheLysAspGluAspGlyAsnIlePheAlaIleGlyAlaGlnAspMetLysCysVal 112
OY 484 GTGATGGCATCTACTGAGGCTTGAGAGCTGCTGCTG-----ATCAGAAAGTACATC 534
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 113 GlyValGlnGlnMetGluAlaLeuArpAsnLeuPheAlaGlnGlyIleArpGlnTrp-- 131
OY 535 CCGCGAAGATCTTCTTCATTTCTCTGCGCATGATGAGAGTCAATCAGGG--ACAGGG 591
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 132 ---LysArpThrIleHisLeuValTrpGlyProAspGluGlnIlePheGlyIleAsnGly 150
OY 592 GCTAGAGGATCTCAGCCCTGCTACAGTCAAGGGGCGTCCAGTACGCTTCAATTTGGAC 651
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 151 MetLysGluPheAlaLysThrAspGluPheLysLysLeuAsnLeuGlyPheSerLeuAsp 170
OY 652 GAGGGGGGCTCATCTGGATGATTTTCATCTCAACTCAAGAAAGCCATCGCTTGAT 711
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 171 GluGlyMetProSerAspAspAsp-----ValTyrLysVal 182
OY 712 GCAGTCTCAGAAAGGCTCCATGAACTCATGCTGCAAGTAAACATGATTCAGGCCAC 771
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 183 PheTyrAlaGluArpValAlaIleTrpValIleValThrPheProGlyAsnProGlyHis 202
OY 772 TCTTCAGCTCCTCCAAAGGAGACAGCATTTGCCATCTTCAGCTGCTGTCAGCCGATTG 831
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 203 GlySerGlnPheMetGluAsnThr-----AlaMetGluLysLeu 215
OY 832 GAGCAGACACCAATGCTATCATATTGGAAGGGGACAGTGGTACTGATTGCAAGAA 891
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 216 GluArg----- 217
OY 892 CTGGCAATGATTTCCCTTCCTGTCATATATATCCCTGAGCAACCATGGTATTGAA 951
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 218 -----TyrLeuAlaSerAlaArgLysPheArg 226
OY 952 CCACTTATAGCAGGTTTATGAGAGAAATCC---TTAACAATGCAATATATCAGAGC 1008
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 227 AspGluGlnLysAlaLeuLeuGlnSerAsnProAspLeuThrIleGlyAspVal---Thr 245
OY 1009 ACCAGGCACTCAACATATCAAGAGGAGGTCAGTCAATGTCATCCCGCCAGTGGCC 1068
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 246 ThrLeuAsnValAsnIleValAsnGlyValGlnPheAsnValIleProGluLysPhe 265
OY 1069 CAGGCCACAGTCACTTCGATTCACCTCGACAGACAGTCCAGAGGTCTAGAAATC 1128
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

Db 266 GluAlaPheValAspIleArgLeuThrProSerIleAspPheAsnGluMetArgAsnLys 285
OY 1129 ACGAAGAACATGTGTGGCTGAT-----AACAGAGTCCAGTTCATGTGTTGAGTGCCTT 1182
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 LeuAspGlnTrpValLysAspAlaGlyGluGlyValThrTyrGluPheSerLysHisSer 305
OY 1183 GACCCCTCCCGCTCAGCCCT-----TCTGATGACAAAGGCTTGGGCTACAG-- 1230
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 306 AspLeuLysLeuValThrProHisThrArpAspAspProPheThrValAlaPheGluAsp 325
OY 1231 CTGCTCGGACAGACCGTACAGTCCGCTTCGCGAGGTCAATATCTGCCCACTACT 1290
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 326 SerLeuLysGlnIleLysCysLysPheThrThrGlnVal----- 338
OY 1291 TCTATTGGCAACACAGACAGCCGATTTCTTACAACTCACACATGCGATCTACAGGTTG 1350
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 339 LeuIleGlySerThrAspSerArgIleValArgGluAlaGlyValArgAlaIleAsnPhe 358
OY 1351 TACCCCATCTACATACAGCT-----GAAAGACTTCAAAAGCATTCAT 1392
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 359 SerProLeuIleAsnThrProLeuLeuAlaHisAlaHisAsnGluPhe----- 374
OY 1393 GGAGTCAAGAGAAATC-----TCAGTCCAAAGCTATGACAGCCCAAGTG 1437
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 375 ---LeuAsnGluLysValPheLeuArgGlyIleGluIleTyrGlnThrLeuIle 391

```

Search completed: June 27, 2003, 11:05:36
 Job time : 87.5 secs

us-10-014-896-1.n2p.rsp

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:      June 27, 2003, 10:51:58 ; Search time 28 Seconds
              (without alignments)
              4470.556 Million cell updates/sec
```

Title: US-10-014-896-1

Sequence: 1 atggtcagcggtgcgtttg.....ctcacctgcacaactgtga 1509

Scoring table: BLOSUM62

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

0-/csm2.1/USPO/pwd/US10014896/runat.27062003.104413.0368/fasta_1.167
-DB-5wsp1rpt.40 -OEMF-fastaan -SUFFIX=02p_rsp -MIMMARCH=0.1 -LOPCG=0
-IOBPEXT=0 -UNITS=bits -SPD=1 -MATRIX=blomsmed -TRANS=humad0.cd1
-115r-45 -DOCALLIGN=200 -RFR SCORE=1 -END=1 -MATRIX=100 -RFR MIT=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ppa -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10014896-ECGN_1_1_34 -ETIMAT.27062003.104413.10368 -NCPU=6 -ICPD=3
NO_MMAP -LANG=PERL -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
DEV.TIMESOUT=120 -WARN.TIMESOUT=30 -THREADS=1 -XGAOP=0.5 -FGAOP=6
FGAEXT=7 -YGAOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	length	DB	ID	Description
No.	Score					
1	540	19.6	576	1	CBPS_YEAST	P27614 saccharomyc
2	202	7.3	406	1	ACV1_PIG	P37111 sus scrofa
3	199	7.2	410	1	Y457_METUA	O57899 methanococc
4	185.5	6.7	408	1	ACV1_HUMAN	O03154 homo sapien
5	180	6.5	375	1	DAPE_BUCAI	P57196 buchnera ap
6	170	6.2	377	1	DAPE_HABIN	P44514 haemophilus
7	163.5	5.9	375	1	DAPE_ECOLI	P24176 escherichia
8	160	5.8	878	1	YB9X_YEAST	P38149 saccharomyc
9	149	5.4	470	1	PEPY_LACDL	P45494 lactobacillu
10	146.5	5.3	403	1	YGEY_ECOLI	O46805 escherichia
11	143.5	5.2	381	1	ARGE_BUCAI	P57155 buchnera ap
12	138.5	5.0	383	1	ARGE_ECOLI	P23908 escherichia
13	134.5	4.9	481	1	YFL4_YEAST	P43366 saccharomyc
14	116	4.2	3178	1	YKS9_YEAST	O09624 caenorhabdi
15	115.5	4.2	383	1	HIPO_CAMEL	P45493 campylobact
16	114	4.1	890	1	ATSR_HUMAN	O9upd9 homo sapien
17	113	4.1	421	1	Y236_METUA	O57688 methanococc
18	108	3.9	1231	1	BLM_CAMEL	O18017 caenorhabdi

19	107	3.9	1305	1	RPOC_UDEPA	Q09pvs ureaplasma
20	106.5	3.9	1140	1	YM96_YEAST	Q04893 saccharomye
21	106	3.8	551	1	YGL1_YEAST	P53114 saccharomye
22	105.5	3.8	454	1	PUC02_RHOSU	P95556 rhodovulum
23	105	3.8	552	1	ASO_CUCDPM	P37064 cucurbita p
24	104.5	3.8	643	1	PGT_RAT	Q00910 rattus norv
25	104.5	3.8	1306	1	MSB2_YEAST	P32334 saccharomye
26	104.5	3.8	2472	1	NCR2_MOUSE	Q9w442 mus musculu
27	104	3.8	1502	1	N170_YEAST	P38181 saccharomye
28	103.5	3.7	364	1	DRN2_PIG	Q62855 sus scrofa
29	103.5	3.7	892	1	ATX7_HUMAN	Q15165 homo sapien
30	103.5	3.7	2845	1	APC_MOUSE	Q61315 mus musculu
31	103	3.7	977	1	YAP9_SCHPO	Q09872 schizosacch
32	102.5	3.7	369	1	DAGE_CORGL	Q09284 corynebacte
33	102.5	3.7	1001	1	RR44_YEAST	Q08162 saccharomye
34	102.5	3.7	1429	1	EXP4_DROME	Q07364 drosophila
35	102	3.7	3969	1	HRX_HUMAN	Q03164 homo sapien
36	101.5	3.7	499	1	C719_SOYBN	Q081970 glyicine max
37	101.5	3.7	691	1	PEP1_YEAST	P26570 saccharomye
38	101	3.7	571	1	ILVY_BACST	Q04789 bacillus m
39	101	3.8	1367	1	AMY1_YEAST	P08640 saccharomye
40	100.5	3.6	377	1	NTRB_BRNSR	P10578 bradyrhizob
41	100.5	3.6	861	1	IMB1_YEAST	Q061442 saccharomye
42	100.5	3.6	1609	1	F1G2_YEAST	P26563 saccharomye
43	100	3.6	462	1	TGCE_DROME	P22812 drosophila
44	100	3.6	3381	1	PUCB_BOVIN	P81882 bos taurus
45	99.5	3.6	579	1	ASO_CUCDA	P24792 cucurbita m

ALIGNMENTS

RESULT 1
CBPS_YEAST

ID	CBS#	YEAST	STANDARD;	PRT;	576 AA.
AC	P27614;				
AD	01-AUG-1992 (Rel. 23, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	13-JUN-2002 (Rel. 41, Last annotation update)				
DE	Carboxypeptidase S precursor (BC 3.4.17.4) (YSCS) (GLY-X				
DE	carboxypeptidase).				
GN	CP81 OR CP5 OR YJL172W OR J0510.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=S288c;				
RC	MEDLINE=91224132; PubMed=2026161;				
RA	Spormann D.O., Helm J., Wolf D.H.;				
RT	"Carboxypeptidase yscs: gene structure and function of the vacuolar				
RT	enzyme.";				
RL	Eur. J. Biochem. 197:399-405(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AB320;				
RC	MEDLINE=91243870; PubMed=1709881;				
RA	Bordallo J., Bordallo C., Gascon S., Suarez-Rendueles P.;				
RT	"Molecular cloning and sequencing of genomic DNA encoding yeast				
RT	vacuolar carboxypeptidase yscs.";				
RL	FEBS Lett. 283:27-32(1991).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Obmerster B., Piravandi E., Rinke M., Dondley H.;				
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RN	SUBCELLULAR LOCATION.				
RX	MEDLINE=92235008; PubMed=1569061;				
RA	Spormann D.O., Helm J., Wolf D.H.;				
RT	"Biogenesis of the yeast vacuole (lysosome). The precursor forms of the				
RT	soluble hydrolyase carboxypeptidase yscs are associated with the				
RT	vacuolar membrane.";				
RL	J. Biol. Chem. 267:8021-8029(1992).				

```

RN [1] SIMILARITY TO ARGE/DAPE/ACY1/CPG2/YSCS FAMILY.
RP MEDLINE=94245187; PubMed=8188249;
RX Henkoff S., Henkoff J.G.;
RT "Protein family classification based on searching a database of
RT blocks.";
RL Genomics 19:97-107(1994).
CC -1- FUNCTION: NECESSARY FOR USE OF CERTAIN PEPTIDES AS SOLE NITROGEN
CC SOURCE. MAY ALSO CLEAVE INTRACELLULARLY GENERATED PEPTIDES TO
CC RECYCLE AMINO ACIDS FOR PROTEIN SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: Peptidylglycine + H(2)O -> peptide + glycine.
CC -1- PATHWAY: NITROGEN METABOLISM.
CC -1- SUBUNIT: YSCS IS SYNTHESIZED AS ONE POLYPEPTIDE CHAIN PRECURSOR
CC WHICH AFTER CARBOHYDRATE MODIFICATION IN THE SECRETORY PATHWAY
CC YIELDS TWO ACTIVE PRECURSOR MOLECULES. THE PROTEOLYTICALLY
CC UNPROCESSED FORMS ARE ASSOCIATED WITH THE MEMBRANE, WHEREAS THE
CC MATURE FORMS OF THE ENZYME ARE SOLUBLE.
CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
CC -1- PIM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: X57316; CAA40571.1; -
CC EMBL: X63068; CAA44790.1; -
CC EMBL: X49447; CAA89467.1; -
CC PIR: S16881; S16881.
CC PIR: S16693; S16693.
CC DR MEROPS: M20.002; -.
CC DR SCD: S0003708; CP51.
CC DR InterPro: IPR001261; ARGE_DAPE_CPG2.
CC DR InterPro: IPR002933; Peptidase_M20.
CC DR Pfam: PF01546; Peptidase_M20; 1.
CC DR PROSITE: PS00756; ARGE_DAPE_CPG2.1; 1.
CC DR PROSITE: PS00759; ARGE_DAPE_CPG2.2; 1.
CC KW Hydroxylase; Carboxypeptidase; Glycoprotein; Signal.
CC FT SIGNAL 1 39
CC FT CHAIN 1 39
CC FT CARBOHYD 88 576
CC FT CARBOHYD 176 176
CC FT CARBOHYD 228 228
CC FT CARBOHYD 381 381
CC FT CARBOHYD 525 525
CC FT CONFLICT 387 387
CC FT SEQUENCE 576 AA; 64597 MW; 5CBB536D421B5F70 CRC64;
CC
CC Alignment Scores:
CC Pred. No.: 1.92e-32
CC Score: 540.00
CC Percent Similarity: 47.49%
CC Best Local Similarity: 31.86%
CC Query Match: 19.55%
CC DB: 1
CC
CC US-10-014-896-1 (1-1509) x CBDS_YEAST (1-576)
CC
CC 157 GAGCGCGTGAAGGTCGCATCAGATTCCCAACAGTCACTTTAGCTGTGAGAGTCCAAAT 216
CC ||| ||| |||:||||| ||| |||
CC Db GLUTylaseusera:snlaValArgIleProthValVal -----GlnsApIysAsnPro 109
CC
CC 217 ACACACGCC-----CTGGCTGAGTTCGGAAATACATTCAT 252
CC ||| ||| |||
CC Db AsnProAlaAspAspProAspPheTyrYlYsHisPheTyrGluLeuHisAspTyrPheGlu 129
CC
CC 253 AAGTCTTTCTTCAGAGGGTCACACACAGATTATACAGCATAAATCGTGAAGATAT 312
CC ||| ||| |||:||||| ||| |||
CC Db LysThrPheProAsnIle-----HisYlYsHisLeuYsLeuGluYsValAsnGluLeu 147

```



```

QY 886 CAGCAATGCGCAATGAGTTCCCTCCCTCATATATATCTGAGCAACCCATGCGTA 945
Db 222 -----IleAsnSerIleuAla----- 227
QY 946 TTGGAACACTATATAGCAGGTTTATGAGAGAAATCCCTTAACCAATGCAATATACAG 1005
Db 228 PheArgGluLysGluLysGlnArgLeuInserAsnGlnLeuLysProGlyAlaVal--- 246
QY 1006 ACAACACAGGCGCATCAGCATATTCAAAGCAGGAGGCTCAAGTTCATGTCATCCCGCAGTG 1065
Db 247 ThrSerValAsnLeuThrMetLeuGluGlyAlaValAlaLysValAsnValProAlaThr 266
QY 1066 GCCCAGCGCCACAGTCACTTCGCGATTCACCTCGAGACAGACAGTCCAGAGGCTCTAGAA 1125
Db 267 MetSerAlaCysPheAspPheArgValAlaProAspValAlaPheLysAlaPheGluGlu 286
QY 1126 CTCACGAGAGAC-----ATTGTGCTGATTAACAGATCCAGTCCATGCTGTGAGTGGC 1179
Db 287 GlnLeuGlnSerTrpCysGlnAlaIleGlyGluGlyAlaThrPheGluPheValGlnLys 306
QY 1180 TTGACCCCTCCCTCCGACGCTTCTGATGACAGGCTTGCGCTACGAGTGCCTCGC 1239
Db 307 TrpMetGluThrGlnValThrSerThrAspAspSerAspProTrpTrp----- 322
QY 1240 CAGACCGTACAGTCCCTCTCCCGAGTCAATATTCTGCCCCAGTTACTTCTTATT-- 1296
Db 323 AlaAlaPheSerGlyValPheLysAspMetLysLeuAlaLeuGlnLeuGluLysPro 342
QY 1297 GGCACACAGACAGCCGATCTTACAAACCTCCAGTCCAGTCCATCTACAGTCTACCCC 1356
Db 343 AlaSerThrAspAlaArgLysIleArgAlaAlaIleGlyValProAlaLeuGlyPheSer 362
QY 1357 ATCTACATACAGC 1369
Db 362 MetAsnHisThr 366

RESULT 3
Y457_METUA STANDARD; PRT; 410 AA.
ID Y457_METUA
AC Q57899;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ0457.
GN MJ0457.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sulten G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overhaug R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodde A.,
RA Scott J.L., Geoghegan N.S.M., Melman J.F., Fuhmann J.L., Nguyen D.,
RA Otterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii*.
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: COULD BE A PEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67496; AAB98445.1; -.
DR MEROPS: M20. UNK; -.
DR TIGR: MJ0457; -.
DR InterPro: IPR002933; Peptidase_M20.
DR Pfam: PF01546; Peptidase_M20; 1.
DR Hypothetical protein; Hydrolase; Metalloprotease; Complete proteome.
SQ SEQUENCE 410 AA; 47170 MW; 90E7ADB25339D86 CRC64;

Alignment Scores:
Pred. No.: 3,84e-07 Length: 410
Score: 199.00 Matches: 99
Percent Similarity: 38.08% Conservative: 83
Best Local Similarity: 20.71% Mismatches: 184
Query Match: 7.20% Indels: 112
DB: Gaps: 19

US-10-014-896-1 (1-1509) x Y457_METUA (1-410)
QY 151 ATGAAGAGAGCGGTGAAA-----GGTCCATTCAGATTCAACAGTACTTTAGC 201
Db 4 IlegGluAlaIleLysLeuGlnSerAspLeuIleArgIleAsnSerValAsnProSer 23
QY 202 -----TTCAGAAATGTCATACTACAGCCCGCTGAGTTCGGAAATACATT 249
Db 24 PheGlyGlyLysGlyGlyLys-----GlnLysAlaGluVal 36
QY 250 CATTAAGCTTTCTTCACTACAGTGTGACACACAGCTTATTCAGCATGAAGTGTGAGAG 309
Db 37 LysLysLysLeuMetGluValAlaGlnSerThrAsnIleGluAsnThrLeuLysGlu 56
QY 310 TATAGCCACCTG-----TTACATATCCA--- 333
Db 57 TyrAsnIleIleAspLysThrGlyIleGluArgProAsnIleValPheLysIleAspPhe 76
QY 334 GGCTCGAGCCCGAGCTTCGAGCCCTACCTGATGCTCATCTTATGATGTGCTGCGCC 393
Db 77 GlyArgAspLysThrLeuHis-----IleIleSerHisLeuAspThrValPro--- 92
QY 394 CCGTAAGAAGGC-----TGGAGGTGCCCCCATCTCTGGGTGAGCGTAT 441
Db 93 -----GluGlyAspIleSerLeuThrPheLysThrAsnProLysGluProValIleLysAsp 110
QY 442 GGCTCATCTATGTCGTGGGGGCGACACTGAGCAAGAACTGTGATGGCATTTCTGACG 501
Db 111 GlyLysIleThrGlyArgGlySerGluAspHisLysGlyIleValSerSerLeu 130
QY 502 GCGTTGAGCTTCGCTGATGACAGAAATACCCCGCAGATCTTCTCATTTCTCTG 561
Db 131 LeuLeuLysMetIlePheGluAsnAsnIleGluProLysThrAsnLeuSerLeuIlePhe 150
QY 562 GGCCATGATGAGGAGTCTATCAGGAGGAGGCTCAGAGATCTAGCCCTGCTACAGTCA 621
Db 151 ValSerAspGluGluAspGly-----SerGluThrGlyLeuLysThrLeuLeuAsn 168
QY 622 AGGGCGCTCAGCTAGCCCTTCATGTGAGCAGGAGGGGCTCATCTGATTTCTAT 681
Db 169 -----PheGluAspIleLeuLysLysAspPheLeuIle 180
QY 682 -----CCTAATTCAGAAAGCCCATCGCC---TTGATTCAGCTTCAGAGAGGCTTC 732
Db 181 IleValProAspPheGlyThrProThrGlyGluPheValGluIleGlyLysGlyIle 200
QY 733 ATGAACCTCAGCTGCAAGTAAATATGATCTTACAGGCCATCTTCAGCTCTCCAAAGAG 792
Db 201 LeuThrPheLysPheAsnIleLysGlyLysGlnCysHisLysSerThrProGluAsnGly 220
QY 793 ACAAGCATTTGCGATCTTGAGCTGCTGATGAGCGATTTGAGACAGACCAATGCCATATC 852
Db 221 LeuAsnAlaAspIleValIleAspAsnPheLeuAsn----- 232
QY 853 ATATTGGAAGCGGACAGTGTGATCTGTATTCACCAACTGCAATGAGTTCCCTTC 912

```

```

Db 223 -----GluLeuTyrasnGlyLeuTyrGluTyrPheaspGlu 244
QY 913 CCTGTCAATATATCCGAGACACCCATGGCTATTTCACACTTAAACAGGTTATG 972
Db 245 ILeasnSerIlePheLeuProGluTyrSerThrPheGluProThrIleLeuLysAsnLys 264
QY 973 GAGAGAAATCCCTTAACCAATGCAATATACAGACACACGCGACTCACCATATTCAAA 1032
Db 265 ValGluAsnPro----- 268
QY 1033 GCAGGGGTCAAGTTCATATGTCATCCGCCAGTGGCCAGCCAGTCAACTTCGGATTT 1092
Db 269 -----AsnThrIleProGlyTyrValGluValAlaPheaspCysArgIle 283
QY 1093 CACCTGTGACAGACAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1146
Db 284 LeuProThrTyrLysIleGluGluValLeuGluPheIleAsnLysPheIleLysAsnPhe 303
QY 1147 -----GATPACAGAGTCCAG-----TTCAT 1167
Db 304 GluPheLysLysTyrIleLysHisTyrAspAsnSerIleLysAlaGluIleThrTyrGlu 323
QY 1168 GTGTGAGTGCCTTTGACCCCTCCCTCCAGCCCTTGTATGACAGGCTTGGGCTAC 1227
Db 324 IleLeuLysSerGluAsnPro-----AsnTyrThrAspGluAsnAlaGluIleIle 340
QY 1228 CAGTGTCCGCGACAGCCGATACAGTCCGTCTCCCGAAGTCAATATCTGCCCCAGTT 1287
Db 341 LysGluLeuLysLysAlaIleLysAsnVal-----LeuAsnThrAspAlaLysLeu 357
QY 1288 ACTTCTATGGCAACACAGACAGCCGATCTTTACAAACCTCACACCTGGCATCTACAGG 1347
Db 358 CysGlyMetGlyGlyGlyTyrValAlaAlaPheLeuArg-----TyrLys 372
QY 1348 TTCTACCCCATATACATACAGCTTAAGACTTCAACACCATTCATGATGATTCAG 1407
Db 373 GlyTyrAsnValAlaValTyrPglYIleGlyGluGluThrAlaHisGlnProAsnGluHis 392
QY 1408 ATCTGATCGACAGCTTATGAGACCCAGTGAATTCATCTTGAATTCATTCAG 1461
Db 393 IleLysIleGluAspLeuValLysMetAlaGluValPheTyrGluIleLeuLys 410

RESULT 4
ACYL_HUMAN STANDARD; PRT; 408 AA.
ID 003154;
AC 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Aminoacylase-1 (EC 3.5.1.14) (N-acyl-L-amino-acid amidohydrolase)
DE (ACYL-1).
GN ACYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=93363640; PubMed=8357837;
RA Mitla M., Kato I., Tsunasawa S.;
RT "The nucleotide sequence of human aminoacylase-1."
RL Blochim. Biophys. Acta 1174:201-203(1993).
RN [2]
RP SEQUENCE FROM N.A.
TISSUE=Liver;
MEDLINE=93352474; PubMed=8394326;
RA Cook R.M., Burke B.J., Buchnagen D.L., Minna J.D., Miller Y.E.;
RT "Human aminoacylase-1. Cloning, sequence, and expression analysis of
a chromosome 3p21 gene inactivated in small cell lung cancer."
RL J. Biol. Chem. 268:17010-17017(1993).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RA Iwaki K., Tanaka Y., Ohta T., Fukuda S., Kurimoto M.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Lung, and Skin;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE HYDROLYSIS OF N-ACETYLATED OR
CC N-ACETYLATED AMINO ACIDS (EXCEPT L-ASPARATE).
CC -1- CATALYTIC ACTIVITY: An N-acyl-L-amino acid + H(2)O = a fatty acid
CC anion + an L-amino acid.
CC -1- COFACTOR: ZINC.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasm;c.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; L07548; AAA02852.1; -
CC EMBL; D14524; BAA03397.1; -
CC EMBL; D16307; BAA03814.1; -
CC DR EMBL; BC000545; AAH00545.1; -
CC DR EMBL; BC003023; AAH03023.1; -
CC DR EMBL; BC014112; AAH14112.1; -
CC DR PIR; S35709; S35709.
CC DR MENOPS; M20.973; -.
CC DR Genew; HGNC:177; ACYL.
CC DR MIM; 104620; -.
CC DR InterPro; IPR001261; ARGE_DAPC_CPG2.
CC DR InterPro; IPR002933; Peptidase_M20.
CC DR Pfam; PF01546; Peptidase_M20.
CC DR PROSITE; PS00758; ARGE_DAPC_CPG2_1; 1.
CC DR PROSITE; PS00759; ARGE_DAPC_CPG2_2; 1.
CC DR HydroLase; ZINC.
CC SQ SEQUENCE 408 AA; 45885 MW; 293350CD7759826C CRC64;

Alignment Scores:
Pred. No.: 3.87e-06 Length: 408
Score: 185.50 Matches: 89
Percent Similarity: 39.14% Conservative: 57
Best Local Similarity: 23.86% Mismatches: 158
Query Match: 6.72% Indels: 69
DB: 1 Gaps: 13

US-10-014-896-1 (1-1509) x ACYL_HUMAN (1-408)
QY 319 CTGTCACTATACCAAGCTCGGACCCAGTTCGAGGCTTACTGTATGCTACTTT 378
Db 62 ValLeuThrTyrProGlyThrAsnProThrIleuSerIleLeuAsnSerHisThr 81
QY 379 GATGTGTGCTGCCCCGTGAAGAGCTGGAGGTGCCCATTTCTGTGGTTG---GAG 485
Db 82 AspValValProValPheLysGluHisTyrSerHisAspProPheGluAlaPheLysAsp 101
QY 436 CGTGATGGCGTCATGTATGTCGGGGGACACTGGACGACGACGAAGTCTGTGATGGCAATTA 495
Db 102 SerGluGlyTyrIleTyrAlaArgIleValaGlnAspMetLysCysValSerIleGlnTyr 121
QY 496 CTGACAGCTTGAGACCTCTGCTGTATGACGAAGTACATCCCGAAGATCTTTCATTT 555
Db 122 LeuGluAlaValAlaArgLysLeuLysValGluGlnHisArgPheProArgThrIleHisMet 141
QY 556 TCTCTGGGCGATGATGAGAGTCAATCATGAGGACAGGGGCTCAGAGAGATCTCAGCCCTGCTA 615
Db 142 ThrPheValProAspGluGluValGlyGly-----GCCTTCAATTGTGAGAGGGGGCC 660
QY 616 CAGTCAAGGGGCGCTCCAGCTA-----

```

```

DB 152 ---hisnglymeGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170
OY 661 TTCATCTGGATGATTCATCTCACTTCAAGAGCCATCGCCTGATTCAGATC--- 717
DB 171 PheAlaLeuArgPheGluGlyIleAlaAsn-----ProThrAspAlaPheIleValPhe 187
OY 718 ---TCAGAGAGAGGTTCCATGACCTGCTGCAGAGTAAACATGACTTACAGCCACTCT 774
DB 188 TyrSerGluArgSerProTyrPheValArgValHisSerThrGlyArgProGluHisAla 207
OY 775 TCAGTCTCCCAAGAGAGAGATTCGGATCCTTCAGCTGCTGAGCCGATTCAGAG 834
DB 208 SerArgPheMetGluAspThr-----AlaAlaGluIleValHis 220
OY 835 CAGACACCAATCCATCATATTTGAGAGCGGAGCAGTGTGACTGATTTGACAGCAATG 894
DB 221 Lys-----ValValAsnSerIleLeuAlaPheArg 230
OY 895 GCAATGAGTTTCCTCCCTGCAATATATATCCGAGCAACCCATGCTATTGAGCA 954
DB 231 GluIysGluTyrP-----GlnArgLeuGlnSerAsnProHisLeuIleGly 246
OY 955 CTTATAGCAGGTTTATGAGAGAAATCCCTTAACATGCAATATATCAGGACCCAGC 1014
DB 247 SerVal-----ThrSerVal 251
OY 1015 GCATCACCATTATTAAGCAGGAGTCAATGTCATGTCATCCCTCCAGTGGCCAGCC 1074
DB 252 AsnLeuThrIleValGluGlyIleValAlaIleValIleProAlaIleMetSerAla 271
OY 1075 ACAGTCAATCCCGATTCACCTGACAGAGCAGTGCAGAGAGTCTCAGAACCTCAGAG 1134
DB 272 SerPheAspPheArgValAlaProAspValAspPheIleValPheGluGlnLeuGln 291
OY 1135 AAC-----ATTGTGGCTGATRAACAGACTCCAGTTCATGCTGTGAGTGGCTTGACCC 1188
DB 292 SerTyrCysGlnAlaIleGlyIleValIleValIleValIleValIleValIleValIle 311
OY 1189 CTCCTCCAGCAGCTTCATGATGACAGAGCCTTGGGCTACAGTGTCTCCGACAGCCGTA 1248
DB 312 ProGlnValThrProThrAspSerSerAsnProTyrPheAlaIlePheSerArg----- 329
OY 1249 CAGTCCGCTTCCTCCGAGAGTCAATATCT---GCCCACTGATCTTATTTGGCAACA 1305
DB 330 -----ValCysLysAspMetAsnLeuThrLeuGluIleMetProAlaIleThr 347
OY 1306 GACAGCCGATTCCTTACAAACCTCAGCAGTGCATGCTACAGGTTCTACCCCATACATA 1365
DB 348 AsnAsnArgTyrIleArgAlaValGlyIleValIleProAlaLeuGlyPheSerProMetAlaAsnArg 367
OY 1366 CAGCCT-----GAGACTTCAAAGCCATCCAT 1392
DB 368 ThrProValIleLeuHisAspHisAspIleValIleHis 380

RESULT 5
DAPE_BUCAI STANDARD: PRT: 375 AA.
AC P57196;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Succinyl-diaminopimelate desuccinylase (EC 3.5.1.18) (SDAP).
GN DAPE OR B0095.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbolic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN 11]
RC SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

```

```

RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: N-succinyl-L-2,6-diaminoheptanedioate + H(2)O
CC - succinate + L-2,6-diaminoheptanedioate.
CC -1- COFACTOR: COBAL OR ZINC (BY SIMILARITY).
CC -1- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AP001118; BAB12814.1; -.
DR MEMOPS: M20. DNA: -.
DR InterPro: IPR001261; ARGE_DAPE_CPG2.
DR InterPro: IPR002933; Peptidase_M20.
DR Pfam: PF01546; Peptidase_M20; 1.
DR TIGRPFAMs: TIGR01246; dape_proteo; 1.
DR PROSITE: PS00758; ARGE_DAPE_CPG2_1; FALSE_NEG.
DR PROSITE: PS00759; ARGE_DAPE_CPG2_2; 1.
KW Daminopimelate biosynthesis; Lysine biosynthesis; Hydrolyase; Cobalt;
KW Zinc; Complete proteome.
SQ
SEQUENCE 375 AA; 41289 MW; 28CAB655F950644 CRC64;

Alignment Scores:
Pred. No.: 9.68e-06 Length: 375
Score: 180.00 Matches: 88
Percent Similarity: 38.08% Conservative: 59
Best Local Similarity: 22.80% Mismatches: 151
Query Match: 6.52% Indels: 88
DB: 1 Gaps: 18

US-10-014-896-1 (1-1509) x DAPE_BUCAI (1-375)
OY 301 GTGAGAGATATAGCCACTGCTGTCATCACTCAAGAGCTCGAGCCCATTCAGCCCTAC 360
DB 45 ValAsnAspThrIleValAsnPheThrAlaPheArgIleGlyIleThrIleVal 61
OY 361 CTGCTGATGCTCCTACTTGTGATGTGCT---GCCCTGAGAGAGGCTGGAGGTCGCC 417
DB 62 ThrPheAlaGlyHisThrAspValValProIleGlyIleAspIleValIleValIleValIle 81
OY 418 CCATTCCTGGGCTGAGCAGTATGAGGCTCATGTCATGTCGAGGACACAGTACAGCAAG 477
DB 82 ProPheGlnProValIleValIleValIleValIleValIleValIleValIleValIleValIle 101
OY 478 AACCTGTCGATGAGCATTCAGAGCCCTGAGAGCTCCTGCTGATCAGG-----AGTAC 531
DB 102 GluAlaLeuAlaIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 121
OY 532 ATCCCCCGAAGATCTTCTCATTTCTCTGCGCCATGAGGAGTACAGGATCAGGAGCA--- 588
DB 122 LysGlyIleValGluSerPheLeuIleThr-----SerAspGluGlnSerSerAlaValAsp 139
OY 589 GGGGCTCAGAGAGATTCAGCCCTGCTACAGTCAAGAGGAGGCTCAGTACGCTTC---ATT 645
DB 140 GlyThrIleValIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 159
OY 646 GTGAGAGAGGGGGGCTTC-----ATCTTGATGATTTTCATCTTACCTTACAGAGGCC 659
DB 160 ValGlyGluProSerSerThrAsnIleValIleValIleValIleValIleValIleValIleValIle 175
OY 700 ATCGCCTTGATTCAGTCTCAGAGAGAGGCTTCAGTCAAGCTCAGTCAAGTCAAGTCAAGTCAAGTCAAG 759
DB 176 -----GlyArgArgGlySerIleThrAlaAsnIleThrIleValGly 189
OY 760 ACTTCAGGCACTCTTCAGCTCTCCAAAGAGAGACAGCATTCGACCTTCGAGCTGCT 819

```



```

OY 802 GGCATCTTGCAGCTGCTGTCAGCCGATGAGCAGACCAATGCCATATATTGGA 861
DB 199 -----HisLeuAlaGluAsnProIle----- 205
OY 862 AGCGGAGACAGTGTACTGTATTCGACCACTG-----GCCAAT 900
DB 206 ---HisLysAlaAlaLeuPheLeuGlnGluLeuThrTyrGlnTyrAspLysGlyAsn 224
OY 901 GAGTTCCCTCCCTGTCATATATATATCCGAGCAACCCATGCTATTGTAACCACTTAA 960
DB 225 GTPhePheProPthSerLeuGlnIleAlaAsn----- 236
OY 961 AGCAGGTTTATGAGAGAAATCCCTTAACCAATATATATGAGGACCACGAGGACTC 1020
DB 236 ----- 236
OY 1021 ACCATATTTCAAGAGGGGTC---AGTTCATATGTCATCCCGGAGGCCAGCCACA 1077
DB 237 -----IleHisAlaGlyThrGlySerAsnAsnValIlePro-----AlaGluLeuTyr 252
OY 1078 GTCACTTCGCGATTCACCCCTGAGACAGACAGTC-----CAAGAGGTC 1119
DB 253 IlegInPheAsnLeuAlaTyrCysThrGluValThrAspGluIleIleLysGlnLysVal 272
OY 1120 CTAGAACTCAGGAGAACATGTGCTGATACAGAGTCCAGTTCATGTTGAGAGCC 1179
DB 273 AlaGluMetLeuGlnLysHisAsnLeuLysTyrArgIleGluTyrAsnLeu-----Ser 290
OY 1180 TTTCAGCCCCCTCCGTCAGCCCTTCGATACAGGCTTGCGCTACAGCTCCGCC 1239
DB 291 GlyLysProPheLeuThrLysProGly-----LysLeuLeuAsp 303
OY 1240 CAGACCCGATCAGACCTCCGTCGCGAGATCATATATCTGCCCCAGTT---ACTTCATTT 1296
DB 304 SerIleThrSerAlaIleGluGluThrIleGlyLeuThrProLysAlaGluThrGlyGly 323
OY 1297 GGCACACAGACAGCCGATTTCTTACAACTCAGCAGCTGAGCTCTACCTCC 1356
DB 324 GlyThrSerAspGlyArgPheIleAlaLeuMetLysAlaGluValGluPheGlyPro 343
OY 1357 ATCTCATATACAGCCGAGACTCTCAACGATCCATGAGTGCAGAGAAATCTCAGTC 1416
DB 344 LeuAsn-----SerThrIleHisLysValAsnGluCysValSerVal 357
OY 1417 CAA 1419
DB 358 Gtu 358
RESULT 7
DAPE_ECOLI STRAND: PRT: 375 AA.
AC P24176;
DT 01-MAR-1992 (rel. 21, Created)
DT 01-MAR-1992 (rel. 21, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Succinyl-diaminopimelate desuccinylase (EC 3.5.1.18) (SDAP).
GN DAPE OR MSGB OR B2472.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Bouvier J., Richard C., Higgins W., Bogler O., Stragler S.;
RT Cloning, characterization, and expression of the dape gene of
RL J. Bacteriol. 174:5265-5271(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92355498; PubMed=1644751;
RN B. Georgopoulos C., Aug D.;

```

```

RT "The essential Escherichia coli msgB gene, a multicopy suppressor of
RT a temperature-sensitive allele of the heat shock gene griP, is
RT identical to dape."
RL J. Bacteriol. 174:5258-5264(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9276503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Mada C.,
RA Yamagata S., Horikuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
CC -1- CARBOLACTIC ACTIVITY: N-succinyl-L-2,6-diaminohexanoate + H(2)O
CC -1- succinate + L-2,6-diaminohexanoate.
CC -1- CORFACTOR: COBALT OR ZINC.
CC -1- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -1- MISCELLANEOUS: DAPE/MSGB IS A MULTICOPY SUPPRESSOR OF A
CC TEMPERATURE-SENSITIVE ALLELE OF THE HEAT SHOCK GENE GREPE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: S41760; AB02798.1; -
CC EMBL: X57403; CAA40665.1; -
CC EMBL: AE000334; AAC75525.1; -
CC EMBL: D90875; BAA16346.1; -
CC EMBL: D90876; BAA16350.1; -
CC PIR: A42958; A42958.
CC PIR: A42959; A42959.
CC MEROPS: M20.DNA; -.
CC Ecodene; EGI0208; dape.
CC InterPro: IPR001261; ARGE_DAPE_CPG2.
CC InterPro: IPR002933; Peptidase_M20.
CC Pfam: PF01546; Peptidase_M20; 1.
CC TIGRfam: TIGR01246; dape_protoc; 1.
CC PROSITE: PS00758; ARGE_DAPE_CPG2_1; 1.
CC PROSITE: PS00759; ARGE_DAPE_CPG2_2; 1.
CC Diaminopimelate biosynthesis; Lysine biosynthesis; Hydrolase; Cobalt;
CC Linc; Complete proteome.
CC SEQUENCE 375 AA; 41269 MW; 181FA3D05E88C20E CRC64;
Alignment Scores:
Pred. No.: 0.000163 Length: 375
Score: 163.50 Matches: 85
Percent Similarity: 36.66% Conservative: 50
Best Local Similarity: 23.10% Mismatches: 134
Query Match: 5.92% Indels: 99
DB: 1 Gaps: 18
US-10-014-896-1 (1-1509) x DAPE_ECOLI (1-375)

```

```

OY 373 CACTTGATGTGGTGGTCCCTGAA---GAAGCGTGGAGAGTGGCCCTCTCTGGG 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 HstHspValValProProGlyAspAlaAspArgTrpIleAsnProProheGlyPro 85
OY 430 TTGGACCGTGAAGCGCTATATGTCGGGCGACACTGACAGACAAGCTGTGTATG 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 ThrIleArgAspGlyMetLeuPheGlyArgGlyAlaAlaAspMetGlySerLeuAla 105
OY 490 GCATTTCGACGCGCTTGAGAGCTCCGCTGATCAGG---AAGTACATCCCGGAGAA 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 AlameValValAlaAlaGluArgPheValAlaGlnHisProAsnHisThrGlyArgLeu 125
OY 544 TCTTTCATATTCCTGTGGCCATGATGAGAGTCAAGG---ACAGGGCGCTCAGAG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 AlaPheLeuIleThr-----SerArgGluGluAlaSerAlaHisAsnIleThrValGly 143
OY 601 ATCTCAGCCCTGTACAGTCAGAGGGCGCTCAGACTAGCCCTTC---ATGTGACAGAGGG 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 ValValGluAlaLeuMetAlaArgAsnGluArgLeuAspTyrCysLeuValGlyGluPro 163
OY 658 GCGTTGATC-----TTGATGATTCATTCCTTAAGTTCAGAGCCCATCGCTGAT 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 SerSerIleGluValAlaGlyAspValValLysAsn----- 175
OY 712 GCAGTCTCAGAGAGAGGCTTCATGATGACCTCATGCTCAAGTAAACATGATTCAGGCGAC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 -----GlyArgArgGlySerLeuThrCysAsnLeuThrIleIleGlyValGlnGlyHis 193
OY 772 TCTTCAGCTCTCCCAAGAGACAGACAGATTGCGATCCTTGACGCTGTCAGCCGATTG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 ValAlaIleThrPro----- 199
OY 832 GACGACAGACCAATG-----CCTATCATATTTGGAAGCGGAGAGTGGTACT 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 AlaAspAsnProValHisArgAlaAlaProPheLeu-----AsnGluLeuValAla 216
OY 880 GTATTGACGACACTGCGAAATGAGTTCCCTTCCTGTCATATTAATCTGAGCAACCCA 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 ILeGluTrpAspGlnGlyAsnGluPhe---PheProAla----- 228
OY 940 TGGCTATTGACACTATTAACAGAGTTATGAGAGAAATCCCTTAACCAATGCAATA 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 ----- 228
OY 1000 ATCAGGACACACGACACTACCATATTCAGAGGGGTC---AAGTCAATGTCATC 1056
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 -----TherSerMetGlnIleAlaAsnIleGlnAlaGlyThrGlySerAsnAsnValIle 246
OY 1057 CCCCCAGTGGCCGACGACAGTCACACTTCGCGATTCCACCCGAGACAGATGCAAGAG 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 ProGlyGluLeuPheValGlnPheAsnPheArgPheSer----- 259
OY 1117 GTCTTGAACTGACAGAGAAATGTCGCTGATGATACAGAGATCCAGTCTGTGACT 1176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 ---ThrGluLeuThr-----AspGluMetIleLeuValGlnValIleAla 273
OY 1177 GCCTTTCAGCCCTCCCGCTCAGCCCTTCGTGATGACAGAGCTTGCGCTACAG--- 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 LeuLeuGluLysHisGlnLeuArgTyrThrValAspTrpIlePheSerGlyGlnProPhe 293
OY 1231 -----CTGCTCCGCGACGCGTACAGTCCGCTCTCCCGGAGTCAATATT 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 LeuThrAlaArgGlyLysLeuValAspAlaValAlaAsnAlaValGluHisTyrAsnGlu 313
OY 1276 ACTGCCCCA-----GTACTTCGATGTCGACAGACAGAGCGGATCTTTACAAACCTC 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 IleLysProGlnLeuLeuThrThrGlyGlyThrSerSppLysArgPheIleAlaArgMet 333
OY 1330 ACCACTGGCATCTACAGCTTCAACCCACTACATACAGCCTGAGACTCAACAGCATC 1389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GlyAlaGlnValValGluLeuGlyProValAsn-----AlaThrIle 347

```

```

OY 1390 CATGGAGTCAACAGCAAAATCTCA 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 HstLysIleAsnGluCysValAsn 355

RESULT 8
YB9X_YEAST STANDARD: PRT: 878 AA.
ID YB9X_YEAST
AC P38149;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 98.1 kDa Trp-Asp repeats containing protein in
DE PAFL-MRPL27 intergenic region.
GN YBR281C OR YBR2018.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378722; PubMed=8091861;
RA Holmstrom K., Brandt T., Kalliesoe T.;
RT "The sequence of a 32,420 bp segment located on the right arm of
RT chromosome II from Saccharomyces cerevisiae."
RL Yeast 10:47-62(1994).
CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X76053; CAA53644.1; -.
CC EMBL: Z36150; CAA85245.1; -.
CC PIR: S44543; S44543.
CC PIR: S39137; S39137.
CC MEROPS: M20.0PA; -.
CC SGD: S0000485; YBR281C.
CC InterPro: IPR002933; Peptidase_M20.
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 4.
CC Pfam: PF01546; Peptidase_M20; 1.
CC PRINTS: PR00320; GPROTEINBPT.
CC SMART: SM00320; WD40; 3.
CC PROSITE: PS00678; WD_REPEATS_1; 2.
CC PROSITE: PS00678; WD_REPEATS_2; 2.
CC PROSITE: PS00678; WD_REPEATS_REGION; 1.
CC KMW Hypothetical protein; Repeat; WD repeat.
FT REPEAT 18 57 WD 1.
FT REPEAT 68 107 WD 2.
FT REPEAT 235 274 WD 3.
FT REPEAT 282 322 WD 4.
FT REPEAT 362 405 WD 5.
FT REPEAT 608 651 WD 6.
SQ SEQUENCE 878 AA; 98070 MW; A24C10B4E94C344 CRC64;

Alignment Scores:
Pred. No.: 0.000378 Length: 878
Score: 160.00 Matches: 95
Percent Similarity: 36.98 Conservative: 74
Best Local Similarity: 20.79 Mismatches: 200
Query Match: 5.79 Indels: 88
DB: 1 Gaps: 15

US-10-014-896-1 (1-1509) x YB9X_YEAST (1-878)
OY 34 GTGCTATATGCTGCTTCTTCCCTACGCTGTCAGATGAGGCGGAGGAGG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 IleThrSerLeuLeuSerAlaValProLeuSerSerAsnSerProIleAsnAlaSerSer 416

```

```

QY 94 GAGCATCAAGG-----GCGTCGCAATCCCTTCAGTTGACGAAAGAGAA 141
Db 417 ThrLeuGlnThrAsnLeuTrpAlaIaIaTyGlnSerIleAsnLeuAsnGlnGlu 436
QY 142 CGCGTCCGATGAAGAGCGCGTGAAGAGTCCATCCAGATCCACAGTACTTTAGC 201
Db 437 -----MetLeuAsnThrLeuArgIuLeuIleSerPheGlnThrValSerGlnSer 453
QY 202 TCTGAGAAGTCAATACTACTACAGCCCTGCGTAGTTCGGAATAATACATCAATGCTTT 261
Db 454 LysAspThrThrAsnThrLeuSerLeuArgGlyAlaIaIeTyLeuGlnGlnLeuPhe 473
QY 262 CCTACAGTGTGACACACAGCTTTATTCACAGATGAGAGTGTGAAGAGTACCACTTC 321
Db 474 LeuLysPheGlnIaIaTrpAsn-----SerGlnLeu 483
QY 322 TTCACATATC---CAAGCGTCGAGACCCGAGCTTGACAGCCCTAC----- 360
Db 484 PheProLeuProAspGlyGlyAsnProValIaIaPheAlaIaTyPheGlnGlnGlyLys 503
QY 361 -----CTGCTGATGGCTCCTTGGATGAG 384
Db 504 ValSerGlnValLysGlyAlaLysLysArgIleLeuThrPtyrGlyHisTyraPval 523
QY 385 GTGCGTCCCGCTGAAGAA---GGCTGGAGAGTGGCCCATTCCTGGGTTGAGCGTGAT 441
Db 524 IleSerSerGlyAsnThrPheAsnTrpAsnThrAspProPheThrLeuThrGlyAsn 543
QY 442 GCGGTCACTATGATGGGCGACACACGAGACAGAAAGCTGTCGATGACGACAG 501
Db 544 GlyTyLeuLysGlyArgGlyValSerAspAsnLysGlyProLeuValSerAlaIleHis 563
QY 502 GCGTGGAGCTCTGTGATCAGAGAAATACATCCCGGAGATCTTCTTCATTTCTGTG 561
Db 564 SerValaIaIaTyLeuPheGlnGlnGlyGluLeuValaIaAsnAspValaIaPheLeuValGlu 583
QY 562 GGGCAGATGAGAGATCAGGAGACAGGGGCTCAAGAGATCTCAGCCCTGCTACAGTCA 621
Db 584 GlySerGlnIuIaIeGlySerAlaSerLeuLysGlnValaIaLysGlnLysTyraHisAspIle 603
QY 622 AGGGGCGTCCAGTACCTTCATTTGTCAGAGAGGGGCGTTCATCTTGGATGATTCAT 681
Db 604 IleGlyLysAspIleAspTrpIleLeuLeuSerAsnSerThrTrpValaIaPheGlnGluHis 623
QY 682 CCTAACTTCAAGACCCATCGCTGATTCAGTCTCAGAGAGGGTTCATGAACCTTC 741
Db 624 Pro-----CysLeuAsnTyrglyLeuArgglyAlaIleAsnAla 636
QY 742 ATGCTGCAAGTA-----AACATGACTTACAGGCACTTTCAGCTCCCTCAAGAGAC 795
Db 637 GlnIleLysValaIaTrpSerAspLysProAspGlyHisSerGly-----Leu 651
QY 796 AGCATTTGGATCTTCAGCTGCTGTCAGCCGATGGAGACACACATGCTTACATA 855
Db 652 AsnGlyGlyValaIaTyraSpGluPrometValaIaLeuValaIaLys----- 665
QY 856 TTGGAGAGCGGACAGTGTGATGTCAGCAACCTGGCAATGATGTTCCCTCCCT 915
Db 666 -----IleValSerLysLeuGlnAsnGlnGlnAsnGlnIleMetIlePro 680
QY 916 GTCAATATAATCTCTGAGCAACCCATGATTTGAACCCCTATAAGCAGGTTTATGAG 975
Db 681 ---AspPheTySerProLeuLysAspLeuThrGlnGlnGlyTrpGlnaIaPheGlnLys 699
QY 976 AGAAATCCCTTAACCAATGCAATAATCAGAGCACC----- 1011
Db 700 IleThrGlnLeuAlaAsnIleAspGluAsnThrThrValGlnAspLeuIleThrAsnTrp 719
QY 1012 AGGCACTTCACCATTTCAAGACAGGCTCAAGTTCAAT-----GTC 1053
Db 720 ThrLysProSerLeuSerMetThrThrValLysPheSerGlyProGlyAsnIleThrVal 739

```

```

QY 1054 ATCCCCGAGTGCCCGCCAGGCGACAGTCACTTCGCGATTCACCTTGACAGACAGTCCAA 1113
Db 740 IleProLysSerValaIaThrMetGlyIleSerIleArgLeuValaIaProGlnGlnSerValGlu 759
QY 1114 GAGTCTCTA-----GAATCAGACAGACATGTTGGCTGAT 1149
Db 760 GlnValaIaArgAspLeuLysAlaIaTyLeuGlnGlnGlnSerPheLysGlnLeuLysSerGln 779
QY 1150 AACAGAGTCACTTCATGTCGTCGATGCTTTCAGCCCTCCCGCTCAGCCCTTCGAT 1209
Db 780 AsnHisLeuGlnIaIeLysValaIaLeuAsnGlnIaIaGlnGlyTyraPheGlyAspProThrAsn 799
QY 1210 GACAAAGCGCTTGCTACAGCTGCTCGCCGACGACGTAAGTCCGCTTC 1260
Db 800 His-----AlaTyroIleLeuLysAspGlnIleThrAlaTrp 813

```

RESULT 9

PEPV_LACDL STANDARD; PRT; 470 AA.

```

AC P45494;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Xaa-His dipeptidase (EC 3.4.13.3) (X-His dipeptidase) (Aminoacyl)-
DE histidine dipeptidase) (Carnosinase).
GN PEPV.
OS Lactobacillus delbrueckii (subsp. lactis).
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus
OX NCBI_TaxID=29397;
RN (1)
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=DSM 7290 / WSB7;
RX MEDLINE=95093606; PubMed=7528082;
RA Vongerichten K., Klein J., Matern H., Plapp R.;
RT "Cloning and nucleotide sequence analysis of pepV, a carnosinase gene
RT from Lactobacillus delbrueckii subsp. lactis DSM 7290, and partial
RT characterization of the enzyme."
RL Microbiology 140:2591-2600(1994).
CC -1- FUNCTION: HAS ACTIVITY AGAINST BETA-ALANYL-DIPEPTIDES INCLUDING
CC CARBOSINE (BETA-ALANYL-HISTIDINE).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of xaa-His dipeptides.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: Z31377; CAA83252.1; -.
DR MEROPS: M20.004; -.
DR InterPro: IPR001261; ARGE_DAPE_CPG2.
DR InterPro: IPR002933; Peptidase_M20.
DR Pfam: PF01546; Peptidase_M20; 1.
DR PROSITE: PS00758; ARGE_DAPE_CPG2_1; 1.
DR PROSITE: PS00759; ARGE_DAPE_CPG2_2; 1.
KW Hydrolase; Dipeptidase; Metalloprotease.
SQ SEQUENCE 470 AA; 5190 MW; 488117BAF33E4AB0 CRC64;

```

Alignment Scores:

```

Pred. No.: 0.00208
Score: 149.00 Length: 470
Percent Similarity: 32.41% Matches: 89
Best Local Similarity: 20.46% Conservative: 52
Query Match: 5.39% Mismatches: 146
DB: 1 Indels: 148
Gaps: 16

```

US-10-014-896-1 (1-1509) x PEPV_LACDL (1-470)

OY	133	AAAGAGGAAGCGCGTCCGCGAATGAAAGAGGGG---	CTGAAGAGTCCGATCCAGATCCACA	189
Db	6	LysGIuLeuAlaGIuAlaLysIAspAlaIleuLysAspLeuGIuLeuIleAla	25	
OY	190	GTGACTTTTAAAGCTCTGAGAAAGTCCAAATTA	CTACAGCCCTGGCTAGCTCGAATAATCAATT	249
Db	26	IleAspSerSerGIuAspLeuGIuAsnAlaThrGIuGIuTyrProValGIuLys	43	
OY	250	CATTAAGTCTTTTCTTACAGTGGTGCACACCCAGCTTTATCCAGAT	-----	294
Db	44	-----GIYProValAspAlaMetThrLysPheLeuSerPheAlaLysArgAspGIu	60	
OY	295	-----GAAGTCGGAAGAGTGTATAGCCAGCTGTCTACATTCACAAGGCTGGAGCCC	345	
Db	61	PheAspThrGIuAsnPheAlaAsnTyrAlaGIuArgValAsnThrGIuAlaGIuAspLys	80	
OY	346	AGCTTGCAAGCCCTACCTGCTGATAGGCTCACTTGATGTGGTGGCTGGCCCTGAGAAGGC	405	
Db	81	ArgLeuGIuLys-----IleIleGIuHisMetAspAlaValProAla---	GIuGIuGIuGIu	96
OY	406	TGGAGAGTGGCCCCATTTCTCT---	GGGTGGAGCGATGATGGCGTCATTCATGGTGGGGG	462
Db	97	TyrThrArgAspProPheLysMetGIuIleAspLeuGIuLysArgIleTyrGIuArgGIu	116	
OY	463	ACAATGACAGCAGAAGAATCTGTGATGGCATTA	CTGCAAGGCTTGAGACTCTGCTGATC	522
Db	117	SerAlaAspAspLysGIuProSerLeuThrAlaTyrTyrGIuMetLeuLeuLysGIu	136	
OY	523	AGGAAGTACATCCCGCCAAATCTTCTTCAATTC	TCCTCCGCGCATGATGAGAGATCA	582
Db	137	AlaGIuPheLysProLysLysIleAspPheValLeuGIuThrAsnGIuGIuThrAsn	156	
OY	583	GGGACGGGGCTCAGAGATCTCAGCCCTGACAGTCAAGAGGGGCGTCCAGCTAGCCTTC	642	
Db	157	TrpAlaGIu-----	159	
OY	643	ATTGTGACGAGAGGGGGCTTCATCTTGATGATTTCA	TCTTAATCTTAAGAAAGCCATC	702
Db	160	-----LeaAspTyrTyrLeuLysHisGIuProThrProAsp	171	
OY	703	GCTTTGATTCAGTCTCAGAGAAGGGTTCATGAACTCA	TGATCGTCAAGTAATCAATGACT	762
Db	172	IleValIhe-----	174	
OY	763	TCAGGCCACTTTCAGTCTCTCCAAAGAGACAAGCAT	TGGCATCTTGCAAGCTGCTGC	822
Db	175	-----SerProAspAlaGIu---	179	
OY	823	AGCGGATTTGACGAGACAGCAAAATCCCATCATTA	TTTGGAAGCGGAGACGTGACTGTA	882
Db	180	-----TyrProIleAsnGIuGIuGIuGIuGIuIlePheThrLeu	192	
OY	883	TTTGACGAACTGGCAAAATGAGTTTCCTTCCCTGCAATAT	TAATCTGAGCAACCCATGG	942
Db	193	-----GIuPheSerPhe-----	196	
OY	943	CTATTTGAAACCACTTATTAAGCAGGTTTATGAGAGAAAT	CCCTTAACCAATGCATATATC	1002
Db	197	-----LysAsnAspAspThr	201	
OY	1003	AGGACAGACACGCGCACTACCATTAATTCAAAGAGGGG	CGAAGTTCATGATCCCCCA	1062
Db	202	LysGIuAspTyrValIleuAspLysPheLysAlaGIuAlaThrAsnValThrProGIu	121	
OY	1063	GTGGCCACGCGCACATCACTTCGCGATTCA	CCCTGGACAGACATGCCAAGAGGTCCTA	1122
Db	222	ValThrArgAlaThrIleSer-----GIuProAspLeuGIuAlaValLys	236	
OY	1123	GAAGTCAGAGAAGAATGTTGGCTGATTAACAGAGG	CGAGTTCATGTTGATGAGCCGTTT	1182
Db	237	LeuAlaIyrTyrGIuSerPheIleuAlaAspLysGIuLeuAsp-----GIuSerPhe	252	
OY	1183	GAACCCCTCCCGCTAGCCCTTCTGATGACAAGG	CGTTGGCTTACACAGTCTCCGCCAG	1242

```

Db      253 GLU-----LleasApscluserAlasplieValleuIleGlyln 266
Oy      1243 ACCGTACAGTCCGCTCTTCCCGGAAGTCAATATTACTGCCACAGTTACTTCTATTGGCAAC 1302
Db      267 GLyAlaHsAlaseralAProGlnVal-----Gly 276
Oy      1303 ACAGACGACCGCATTTCTTTTACAAACCTCACACGCTGCATCTACAGATTC----- 1350
Db      277 LysAsnserAlatlrPheLeuAlaLePheLeuAspInlYrAlaPheAlaGlyArgAsp 296
Oy      1351 -----TACCCCATCTACATACAGCTGAGACTTC 1380
Db      297 LysAsnPheLeuHsPheLeuAlaGlyValGlnHsIsgLysAsp 311

RESULT 10
YGEY_ECOLI STANDARD: PRT: 403 AA.
ID YGEY_ECOLI
AC Q46805:
DT 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ygeY.
YGEY OR B2872 OR Z4211 OR EC53745.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=92378503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E., Potamousis K.,
RA Apodaca J., Anahtarman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shimagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC - - - - - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcements
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 028375; AAA83053.1; -.

```


NP SEQUENCE FROM N.A.
 RP [2]
 RC STRAIN-S288C / AB972;
 RX MEDLINE=96287652; PubMed=8686379;
 RA Ekl T., Naitou M., Hagiwara H., Ozawa M., Sasamura S.-I.,
 RA Sasamura M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
 RT "Analysis of a 36.2 kb DNA sequence including the right telomere of
 chromosome VI from *Saccharomyces cerevisiae*.";
 RL Yeast 12:149-167(1996).
 CC -1- SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.
 CC
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC
 DR EMBL, D50617; BAA09283.1; -.
 DR SGD; S0001940; YFR044C.
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; 1.
 KW Hypothetical protein.
 SO SEQUENCE 461 AA; 52871 MW; 3E53773A945F5EBC CRC64;

PT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Alignment Scores:

pred. No.: 1.01 Length: 3178
Score: 116.00 Matches: 100
Percent Similarity: 38.31% Conservative: 59
Best Local Similarity: 24.10% Mismatches: 188
Query Match: 4.20% Indels: 69
DB: 1 Gaps: 14

US-10-014-896-1 (1-1509) x YS89_CAEEL (1-3178)

QY 331 GCCCTGAGAGAGGCTGGAGAGGCTCCCATTTCTGCTGGTTGAGCCT--GATGGCCTC 447
DB 159 AAlaValGlnAlaLysTyrGluValCysTyrAspAspIleLeuAspArgCysAspGlySer 178
QY 448 ATCTATGGTGGGGGACACGACGACAGAACTCTGTGATGGCATTACTACAGGCTTGG 507
DB 179 LeuTrp-----TrrleuGlnValGlyGlyAsnGlnMetAlaLeuLeuGlyTyrArg 195
QY 508 GAGCTCTGCTGATCAGAGATACATCCCGAAGATCTTCTTCATTTCTGCGGCAT 567
DB 196 Glu-LysCysGlnSerGlyGlnLeuAsnGlnGlyTyrAlaArgArgMetCysArgArg 215
QY 568 GATGAGAGATCATTCAGGAGGAGGAGGAGATCTGAGCCTGCTACAGTCAAGGCG 627
DB 215 CTyrArgSerGlnLysSerThrAlaIleSerAspSerGlnGlyValTyrTyrAspGly 235
QY 628 GTCCAGTACGCTTCATTTGAGACAGGAGGCTTCATCTGATGATTCATCTTCTTAC 687
DB 235 nval-----LeuLysGlyValAlaArgAlaLysGlnPheSerMetArgTyr 249
QY 688 TTCAGAGAGCCATGCTGATTCAGAGTTCAGAGAGGCTTCATGATCACTGATCTG 747
DB 249 rSerGlySerProThr-----LeuArgArgMetLysArgAspIleGlyAspAsnThrC 267
QY 748 CA-----GTAACATGAC 761
DB 267 sAspTyrThrIleGlnSerThrSerThrThrThrProThrThrThrThrValThr 287
QY 762 TTCAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
DB 287 rSerThrValThrSerThrThrThrValProThrSerThrSerThrValThrThrAla 307
QY 822 CAGCGGATTTGAGACAGACCA-----ATGCTCATATATTTGAAAGGCGAGATGTG 875
DB 307 tSerThrSerThrSerThrProSerThrSerThrThrIleGlnSerThrSerThrThr 327
QY 876 GACT-----GTATTTGAGCAACGCGCAATGA 902
DB 327 eThrSerThrAlaSerThrSerThrSerThrSerThrThrThrGlnGlnSerSerThr 347
QY 903 GTTT-----CCCTTCCCTGCTCATATATATATATATATATATATATATATATAT 953
DB 347 rIleThrSerSerProSerSerThrThrLeuSerThrSerThrIleProThrThrThrPr 367
QY 954 ACTTTAAGCAGCTTATGAGAGAAACCTTAAACATCAACATCAATCAATCAATCAAT 1002
DB 367 oLiuIleThrSerThrLeuSerSerLeuPro--AspAsnAlaIleCysSerThrLeuAs 386
QY 1003 -AGAGCAACAGGAGCTACCATCATATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1061
DB 386 pGluThrThrThrThrThrThr-ThrPheThrThrThrMetLeuThrSerThrThrGln 406
QY 1062 AGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1121
DB 406 lProSerThrSerThrThrThrThrGlnValThrSerThrSerSerThrValThrThr 426
QY 1122 AGAAGTCAAGAGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1166
DB 426 hGlnProThrThrThrThrThrThrThrSerThrThrAlaSerThrSerThrThrGln 446

QY 1167 TGTGTAGTGCCTTTGACCCCTCCCTGACGCTTCTGATGACAGGCTTGGGCTTA 1226
DB 446 hrSerThrValThrThrSerProSerThrSerProValThrSerThrValThrSer 466
QY 1227 CCAGGTGCTGCGGACGAGACCGGACGAGGCTGCTCCGGAAGTCAATA--TTACTGCCCC 1283
DB 466 erSerSerSerThrThrValThrThrProThrSerThrThrGlnGlnSerThrSer 486
QY 1284 AGTACTCTGATTTGGACAGACAGACGAGGATCTTTACAACTGACACGATGATCA 1343
DB 486 erSerThrValThrThrSerThrThrThrAlaProSerThrSerThrThrGlnGln 506
QY 1344 CAGTTTCACCCCTACATACAGCCTGTAAGACTTCAACGATCATGATGATGATGAT 1403
DB 506 erSerSerThrProSerSer-----ThrAlaSerSerValSerSerThr 522
QY 1404 GAAATCTCACTCCAGGCTTACAGACCAAGTGAATGATGATGATGATGATGATGAT 1463
DB 522 laseSerThr-----GlnSerSerThrSerThrGlnGln 534
QY 1464 TGCTGACAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1504
DB 534 erSerThrThrThrLysSerGlnThrThrThrSerSer 547

RESULT 15
HIPO_CAMJE STANDARD; PRT: 383 AA.
AC P45493: O9PNV4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hippurate hydrolase (EC 3.5.1.32) (benzoylglycine amidohydrolase)
DE (Hippuricase).
GN HIPO OR CJO985C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli.";
RL J. Bacteriol. 177:2396-2402(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jorgensen K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vleet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC 1- CATALYTIC ACTIVITY: Hippurate + H(2)O -> benzoate + glycine.
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M40.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC EMBL: Z36940; CAA85396.1; -
CC EMBL: AL139076; CAB73241.1; -
CC MEROPS: M40.UMW; -.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 10:46:38 ; Search time 125.5 Seconds

(without alignments)
4954.982 Million cell updates/sec

Title: US-10-014-896-1

Perfect score: 2762
Sequence: 1 atgctcagcggtgctgtt.....ctcactgcacacactgtga 1509

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ .n2p.model -DEV=xlp
-O=cpn2.1/USPTO.spool/US10014896/runat.27062003.104414.10381/app.query.fasta_1.1671
-DB=SPRMBL.21 -OFRM=fasten -SUFIX=n2p.rspt -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10014896 -CGCN.1.1.172 -runat.27062003.104414.10381 -NCPV=6 -ICPV=3
-NO_MAMP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPRMBL.21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriophage.*
17: sp_archaeap.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	65.7	361	4	Q96DM4 homo sapien

2	1261.5	45.7	340	11	Q8R117	Q8R117 mus musculu
3	781	28.3	510	16	Q8X517	Q8X517 raietonia s
4	639.5	23.2	488	2	Q8RT11	Q8RT11 uncultured
5	591	19.4	471	16	Q9AA70	Q9AA70 caulobacter
6	531.5	19.2	596	3	Q13968	Q13968 schizosacch
7	434	15.7	469	2	Q8RNM5	Q8RNM5 legionella
8	361.5	13.1	443	16	Q9ZB17	Q9ZB17 streptomyce
9	341.5	12.4	474	16	Q9A365	Q9A365 caulobacter
10	330.5	12.0	467	16	Q9CC46	Q9CC46 mycobacteri
11	318	11.5	441	2	Q93H22	Q93H22 streptomyce
12	309.5	11.2	448	16	Q06234	Q06234 mycobacteri
13	300	10.9	445	16	Q9AD91	Q9AD91 streptomyce
14	276	10.0	442	16	Q93R29	Q93R29 streptomyce
15	253.5	9.2	401	5	Q9VCQ9	Q9VCQ9 drosophila
16	232.5	8.4	382	17	Q970S3	Q970S3 sulfolobus
17	229.5	8.3	397	5	Q17899	Q17899 caenorhabdi
18	228	8.3	400	5	Q9VCR0	Q9VCR0 drosophila
19	227.5	8.2	402	5	Q8T490	Q8T490 drosophila
20	225.5	8.2	378	16	Q92E19	Q92E19 listeria in
21	224.5	8.1	408	5	Q9VCRL	Q9VCRL drosophila
22	223	8.1	410	17	Q97Y12	Q97Y12 sulfolobus
23	217	7.9	340	5	Q9VC08	Q9VC08 drosophila
24	214.5	7.8	382	16	Q92EY0	Q92EY0 rickettsia
25	213.5	7.7	455	17	Q59016	Q59016 pyrococcus
26	212	7.7	438	10	Q9LPE9	Q9LPE9 arabidopsis
27	209	7.6	435	10	Q9CV68	Q9CV68 arabidopsis
28	207.5	7.5	359	5	Q9VCR2	Q9VCR2 drosophila
29	207.5	7.5	399	5	Q17898	Q17898 caenorhabdi
30	205.5	7.4	397	16	Q8XWY5	Q8XWY5 raietonia s
31	205	7.4	388	16	Q91A45	Q91A45 pseudomonas
32	204	7.4	408	11	Q9CR15	Q9CR15 mus musculu
33	203	7.3	379	16	Q9ZEV0	Q9ZEV0 listeria mb
34	201.5	7.3	397	17	Q8ZVD7	Q8ZVD7 pyrobacul
35	200.5	7.3	428	5	Q17686	Q17686 caenorhabdi
36	197.5	7.2	383	16	Q9ZC93	Q9ZC93 rickettsia
37	197.5	7.2	401	5	Q917K3	Q917K3 drosophila
38	197	7.1	374	16	Q92Y75	Q92Y75 rhizobium m
39	197	7.1	379	2	Q9EXF4	Q9EXF4 listeria mo
40	196.5	7.1	394	5	Q17900	Q17900 caenorhabdi
41	196.5	7.1	407	16	Q9S8N6	Q9S8N6 staphylococ
42	195.5	7.1	377	16	Q9K052	Q9K052 vibrio chol
43	195	7.1	457	16	Q9ZT10	Q9ZT10 streptococ
44	194	7.0	388	16	Q9ZMM0	Q9ZMM0 helicobacte
45	194	7.0	408	11	Q99JW2	Q99JW2 mus musculu

ALIGNMENTS

RESULT 1
Q96DM4 PRELIMINARY; PRT; 361 AA.
ID Q96DM4;
AC Q96DM4;
DC 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DR 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE CDNA FLJ32569 f1s, clone SPLEN2000134, weakly similar to
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaitsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
EMBL AK057131; BAB/1368.1; .

DR InterPro: IPR002933; Peptidase_M20.
 DR Pfam: PF01546; Peptidase_M20; 1.
 SQ SEQUENCE 361 AA; 39560 MW; 367DECB0991F314C CRC64;

Alignment Scores:

Pred. No.:	2,04e-144	Length:	361
Score:	1816.00	Matches:	359
Percent Similarity:	93.73%	Conservative:	0
Best Local Similarity:	93.73%	Mismatches:	2
Query Match:	65.75%	Indels:	22
DB:	4	Gaps:	1

US-10-014-896-1 (1-1509) x Q96DK4 (1-361)

```

QY 1 ATGGCTCAGCGGCGGTTGGCTGCTGCGCCCTGGTGGTATGCTGCTTATGTTTCCCT 60
DB 1 MetAlaGlnArgCysValcysValIleuAlaIleuValAlaMetIleuIleuValPhePro 20
QY 61 ACCGCTCCAGATGATGATGGCCCGAGAGCGGGAGCATCAAGCGCGTGGCGAATCCCT 120
DB 21 ThrValSerArgSerMetGlyProArgSerGlyIleuIleuIleuValAlaSerArgIlePro 40
QY 121 TCTCAGTTCAGCAAGAGAGAGCGGCGCATGAAGAGCGCGTGAAGAGCGCGCATCCAG 180
DB 41 SerGlnPheSerIlysgluIuArgValAlaMetIlysgluAlaIleuIlysglyAlaIleGln 60
QY 181 ATTCACACAGTGAAGTCTAGAGAGTCAATCAATCAAGCCCTGAGTGGTGGATGGGA 240
DB 61 IleProThrValIlePheSerSerGlyIlySerSerThrThrAlaIleuAlaGluPheGly 80
QY 241 AATATCATTCATAAGCTTCTCCACAGTGGTCAAGCAGCCGTTTATCCAGCATGAAGTC 300
DB 81 LysTyrIleHisIlyValPheProThrValValSerThrPheIleGlnHisIlyVal 100
QY 301 GTGGAAGATATACCCCACTGTCATATCCCAAGGCTCGAGACCCGCTGAGCCCTAC 360
DB 101 ValGluIuIySerHisIleuPheThrIleGlnIlySerPheProSerIleuGlnProTyr 120
QY 361 CTGCGATGCGCTCACTTGTATGTCGTCGCTGCGCCCTGAGAGAGCGTGGAGTCCCA 420
DB 121 LeuIleuMetAlaHisPheAspValValProAlaProGluIuIyTTPGluValProPro 140
QY 421 TTTCTGGGTTGGAGCGTGAAGGCGCATATGATGGTGGGCGACATCGAGCAAGAAC 480
DB 141 PheSerIlyLeuGluIuArgAspGlyValIleIyTyrIlePheIleuAspAspIyAsn 160
QY 481 TCTGTGATGCGATTAAGTCAAGCGCTTGGAGCTCTGCTGATCAAGAGTATCCCGCA 540
DB 161 SerValMetAlaIleuGlnAlaIleuGluIleuIleuIleuIleuIleuIleuIleu 180
QY 541 AGATCTTCTTCATTTCTCTGGGCGCATGATGAGAGATCATCAGGAGCGGCTCAGAG 600
DB 181 ArgSerPhePheIleSerIleuGlnHisAspGluIuIySerSerGlyThrGlyAlaGlnArg 200
QY 601 ATTCAGCCCTGCTAGAGTCAAGGGGCGTCAAGTGGCTTATGTTGGACGAGGGGCG 660
DB 201 IleSerAlaIleuGlnIleuArgGlyValGlnIleuAlaPheIleValAspGluIy 220
QY 661 TTCATCTTGGATGATTTCACTTCAAGAGCGCCATCCGCTTGGATGGAGTCACTCA 720
DB 221 PheIleuAspAspPheIleProAsnPheIlyIyProIleAlaIleuIleAlaValSer 240
QY 721 GAGAAGGGTTCATGAAGCTCAAGTGGCAAGTAACATGATGATCAAGCGCACTTCAAG 780
DB 241 GluIlysglySerMetAsnIleuGlnIleuValIleuMetThrSerGlnHisSerSerAla 260
QY 781 CCTCAAGAGAGACATGGCATTCCTTGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 261 ProProIySerIlyThrSerIleGlyIleuAlaAlaIleuValSerArgIleuGluIy 280
QY 841 CCAATGCGTATCATTTTGGAGCGGAGACAGTGGTGAATGATTTGGAGAGACATGCAAT 900
DB 281 ProMetProIleIlePheGlySerGlyThrValValIleuGlnIleuAlaAsn 300

```

```

QY 901 GAGTTCOCCTTCCCTGTCATATATATCTGAGCAACCCATGGCTATTGAAACCACTTATA 960
DB 300 -----
QY 961 AGCAGGTTTATGAGAGAAATCCCTTACCAATGCAATATGAGACACAGCGGCACTC 1020
DB 301 Glu-ValTlyGlyIlyIySerIleuAlaGlnCysAsnIleuAsnIleuIleuIleuIleu 320
QY 1021 ACCATATTCAGACAGCGGCTCAAGTTCATGTCATCCCCCAGTGGCCGAGCGCAAGT 1080
DB 320 IHisIleGlnSerArgGlyIleuValGlnCysHisIleProProSerGlyProGlyHisSerG 340
QY 1081 AACTCCGATTCACCCCTGGAGAGACAGTCAAGAGTCTTGAAGTCAAGCAAGCAAT 1140
DB 340 ILeuProAspSerProThrPheAspSerProArgIlyProArgThrIleGluIleuIle 360
QY 1141 GTGGC 1145
DB 360 ysgly 361

```

RESULT 2

```

ID 08R117 PRELIMINARY; PRT; 340 AA.
AC 08R117;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DE 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 37.8 kDa protein (fragment).
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025830; AAL25830.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 340 AA; 37800 MW; 585321EE72EPAD2E CRC64;

```

Alignment Scores:

Pred. No.:	1.09e-97	Length:	340
Score:	1261.50	Matches:	235
Percent Similarity:	85.29%	Conservative:	55
Best Local Similarity:	69.12%	Mismatches:	49
Query Match:	45.67%	Indels:	1
DB:	11	Gaps:	1

US-10-014-896-1 (1-1509) x Q8R117 (1-340)

```

QY 490 GCATTACTGACAGCGCTTGGAGCTGCTGATCAGAAATACATCCCGAAGATCTTC 549
DB 1 AlaiIleuHisAlaIleuGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 20
QY 550 TTCATCTTCCGCGGCGATGATGAGAGTCAAGG---AAGGGCTCAAGATGCTCA 606
DB 21 PheIleuAlaIleuGlnHisAspIleuIleuValSerGlyIlysglyAlaGlnIlyIleSer 40
QY 607 GCCCTGTAAGTCAAGTCAAGGGGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 666
DB 41 AlaIleuGlnAlaIleuArgGlyValGlnIleuAlaPheIleuValAspGluIySerPheIle 60
QY 667 TTGATGATTTCAATCTTAAGTCAAGAGCGCATGCGCTTATGATGAGTCAAGTCAAGAG 726
DB 61 IeugluIyPheIleProAsnIleuGluIyProValAlaMetIleSerValThrGluIy 80
QY 727 GGTCCATGAGACCCATATCTGCAAGTAACATGACTTCAAGCCACTTCAAGCTCTCA 786
DB 81 GlyAlaIleuAspIleuMetIleuGlnValAlaIleuMetThrProGlyIlyHisSerSerAlaProPro 100

```

QY	787	AAGGAGACAGCAAGTTGGCAATCCCTTTCAGAGCGTCGTGACGGCCGATTTGAGGACAGACCCAAATG	846
Db	101	LyglLwHsrIleerIleeglyIleleuSerIaIaIaValSerIyIeugIugIlnItrPromet	120
QY	847	CCATATCATATTTTGGAAAGCGGAGACAGTGGTACTGTATGTCAGCAACATGGCAAAATGAGTTT	906
Db	121	ProasmethenehlygllyglIProleuIyLysIthMetIySleuIeIaIaSnIgluPhe	140
QY	907	CCCTTCCCTTCATATATATATCCCTGAGACACCATGGCTATTTTGAACCAATTTATAGACAG	966
Db	141	SenPheProIleasnIleValIleuIaIySnIleuItrPheunPheIaSnProIleValSerIarg	160
QY	967	TTTATGAGAGAAATCCCTTAACCAATGACATATATTCAGACACACAGGACACTACACATA	1026
Db	161	IlleethcIuIryuSnProIleIthIaSnIaIeUValIygrIhItrIhIaIeIthMet	180
QY	1027	TTCCAAAGCAGGGCTCAAGTTCCATGTGCATGCCCCAGCTGGCCAGGCCACAGTCACTTC	1086
Db	181	PheSnIaIcIyIleIyIValIaSnIValIleIroProIleuIaIcIaIaIhIaIhIleIaSnIyS	200
QY	1087	CGGATTCACCCCTGGACAGACAGTCCACAGAGGCTGTACAGTCCAGCAAGAAATATGTGGCT	1146
Db	201	ArgIleIhIaSProSerGIlnIhIaSnIValIeUgIlnIleUgIlnIValIySnIhIaIaIa	220
QY	1147	GATTAACAGAGTCCAGTTCCATGTGTGTAGTCCCTTTGACCCCTCCCGCTGAGCCCTTCT	1206
Db	221	AspAspIryValGIlnIeUHIaSnIValIeUdIySerPheGIuProIleuProIleSerProSer	240
QY	1207	GATGACAAAGCGCTGGGCTTACCACTGTCTCCGCCAGACCCGAAACAGTCCGTTCGCCGGA	1266
Db	241	AspAspGIlnIaIeMetGIyIryGIlnIeUeUgIlnIuItrIleIaIySerValIpheProGIu	260
QY	1267	GTCATATATTCATGGCCCAAGTACTCTTATATGGCAACACAGACAGCCGATTTCTTACAAAC	1326
Db	261	ValAspIleValIaIProGIyIleCysIleIaIaSnIhIaSnIhIaIryIaIhIaIaIaSnIy	280
QY	1327	CTCACACATGGCATCTACAGTTCTTACCCCACTATACATACAGCCCTGAGAGTATCAAAAGC	1386
Db	281	IleIhIaSnIyMeIyIaIryPheIaSnIroIleuProIleuAaSnProGIaIaSnIaSPheSerGIy	300
QY	1387	ATCCATAGAGATCAACGAGAAATATCTCAAGCTTGTGAGACCCAGTCAATTCATC	1448
Db	301	ValIhIaSnIyIleAaSnIyIyValSerIaIeIaIaSnIyryGIlnAaSnGIlnIaIyIySerIeIe	320
QY	1447	TTTGAGTTGATTCAGATGCTGTCAGACACAGACAGGACCAAGTTTCTCACACTGCACAAATG	1506
Db	321	PheGIuPheIleGIlnAaSnIaIaSPheIryIyGIuGIuProValProIhIaIeUHIaIeUgIleu	340
RESULT 3			
ID	08XS17	PRELIMINARY	PRT: 510 AA.
AC	08XS17		
DT	01-MAR-2002	(TREMBLrel. 20, Created)	
DR	01-MAR-2002	(TREMBLrel. 20, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Hypoethetical transmembrane protein Rsp0487.		
GN	RSP0487	OR R500365.	
OS	Ralstonia solanacearum (Pseudomonas solanacearum).		
OC	Plasmid megaplasmid.		
OC	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;		
OC	Ralstonia.		
OX	NCBI_TaxID=305;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=GM1100;		
RX	MEDLINE=21681879; PubMed=11823852;		
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,		
RA	Arlat M., Billaut A., Brotier P., Camus J.C., Catolico L.,		
RA	Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,		
RA	Gaspin C., Layte M., Moisan A., Robert C., Saurin W., Schlex T.,		
RA	Stigler P., Thebaud P., Whalen M., Wincker P., Levy M.,		
RA	Weissenbach J., Boucher C.A.;		
RT	*Genome sequence of the plant pathogen Ralstonia solanacearum.*;		

[illegible]


```
QY 1273 ATTACGCCCACTTACTCTATTGCAACAGACAGCCGATCTTTACAAACCTGACC 1332
DB      |||
QY 420 LeuAlaIleProSerLeuEnglAlaThrThrAspThrArgHisValValAsnLeuAla 439
DB      |||
QY 1333 ACTGGCATCTACAGGTTCTACCCCATCTACATACAGCCTGAGACCTTCAACGCATCAT 1392
DB      |||
QY 440 LysAspGlnTyrArgPheHisGlyAsnSerIleAspAlaSerGlnAlaArgSerValHis 459
DB      |||
QY 1393 GGAGTCAGCAGAGAAAATCTCAGTCCAGCCTATGACGCCAAGTGAATTCATTCTTGAG 1452
DB      |||
QY 460 GlyThrAsnGluTyrIleSerGluArgSerTyrAsnAsnAlaIleAlaValAlaArgGly 479
DB      |||
QY 1453 TTGATTCAGAAATGCT 1467
DB      |||
DB 480 MetLeuIysGlyAla 484

RESULT 5
Q9AA70 PRELIMINARY; PRT; 471 AA.
ID Q9AA70;
AC Q9AA70;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE M20/M25/M40 family peptidase.
GN CC0736.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MBLDLINE=2113698; PubMed=11259647;
RA Nielsen W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Ely B.,
RA Kolonay J.F., Smit J., Craven M.B., Knout H., Shetty J., Berry K.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005748; MAK2721.1; -.
DR TIGR: CC0736;
DR InterPro: IPR002933; Peptidase_M20.
DR Pfam: PF01546; Peptidase_M20; 1.
KW Complete proteome.
SQ SEQUENCE 471 AA; 49579 MW; 47B67159BBECDEF CRC64;

Alignment Scores:
Pred. No.: 3,99e-41 Length: 471
Score: 591.00 Matches: 151
Percent Similarity: 52.21% Conservative: 85
Best Local Similarity: 33.41% Mismatches: 184
Query Match: 21.40% Indels: 32
DB: 16 Gaps: 12

US-10-014-896-1 (1-1509) x Q9AA70 (1-471)
QY 157 GAGGCGCTGAAGGTCATCGATTCACAGAGTACTTTAGCTGTGAGAGTCCAT 216
DB      |||
QY 34 GlnHisLeuAlaGluAlaIleArgPheGlnThrIleSerHisGlnAsnAlaGlyAlaAsp 53
DB      |||
QY 217 ACTAGACCCCTGGCTGAGTTGCGAAA-----TACATTGATAAAGCTTTCTTACA 267
DB      |||
QY 54 GluPro-----AlaGluTrpAspIleuHisAlaTrpLeuGlnThrTyrProGln 71
DB      |||
QY 268 GTGTGACGACACGCTTATTCACGATTCAGAGTGTGAGAGATTAACGACGCTGTAC 327
DB      |||
QY 72 -----AlaHisLysAlaMetThrArgGluValAlaGluIleGlyLeuValTyrThr 89
DB      |||
QY 328 ATCCAGGCTCGACCCGAGCTTCGACGCTTACCTGCTGATGCTCACTTGTGATGTG 387
DB      |||
```

```
DB 90 TrpThrGlySerAsnProAlaLeuAlaProIleValLeuMetAlaHisGlnAspValAl 109
QY 388 -----CTGCCCCGTAAGAGAGGCTGGAGAGTCCCGCATCTGTGGTGGACGGT 438
DB      |||
QY 110 ProValThrProGlySerGluGlyGlnTrpThrHisProPheAlaGluValValAla 129
DB      |||
QY 439 GATGGCGTCATCTATGTCGGGGGACACTGGACGACGAAAGTCTGTGATGGCATCTAG 498
DB      |||
QY 130 AspGlyLysValTrpGlyArgGlyAlaIleAspAspLysGlySerLeuValThrIlePhe 149
DB      |||
QY 439 CAGGCGCTGGAGACTCTCTGTCATCGAAGTACATCCCGCAAGATCTTCTGATTTCT 558
DB      |||
QY 150 GluAlaLeuGlnSerValAlaAlaGlyGlyIlePheLysProValAlaGlyThrValIle 169
DB      |||
QY 559 CTGGGCGCATGATGAGAGCTCATCAGGACAGGGGCTCAGAGATCTCAGCCCTGACG 618
DB      |||
QY 170 SerGlnHisAspGluGluValArgGlyGlyGluAlaGlnAlaAlaIleLeuLeu 189
DB      |||
QY 619 TCAGGGGCGTCCAGTACGCTTCATGTGTGACGAGGGGGCTTCATCTGGATGATTC 678
DB      |||
QY 190 SerArgAsnIleLysAlaGlnPheValLeuAspGluGlyMetAlaValAlaAspHis 209
DB      |||
QY 679 ATTCCTAATTCAGAGAGCCCATCGCTTATGTCAGTGTGACGAGAGGTTCCATGAC 738
DB      |||
QY 210 ---ProValThrAsnGluProAlaAlaIleIleGlyValAlaGlyGlyTyrAlaThr 228
DB      |||
QY 739 CTCATGCTGCAAGTAACATGACTTCAGGCGCATCTTCAGCTCTCCAAAGAG---ACA 795
DB      |||
QY 229 LeuLysValThrAlaProAlaValAlaGlyHisSerSerAlaProProLysAspGly 248
DB      |||
QY 796 AGCATTTGGATCTCTGAGCTGCTGTCAAGCCGATTTGAGACGACGACGCAATGCTAT 855
DB      |||
QY 249 GlyValValThrLeuSerLysAlaGlnAlaIleHisAspAsnProPheTrpLeuLys 268
DB      |||
QY 856 TTTCGAGGCGGACAGTGTGACTGATTCATGTCAGCAATGCGCAAGAGATTTCCCTCC 915
DB      |||
QY 269 Phe---GlnGlyProGlyAlaAspMetLeuLysAlaIleSerProHisAlaSerProAl 287
DB      |||
QY 916 GTCATATATATCTGACGACACCCATGCTATTGACCACTTATAGCAGTTATGAG 975
DB      |||
QY 288 ValLysValPheAlaAlaLysAsnThrTrpLeuPheSerSerLeuLeuValAlaThrAla 307
DB      |||
QY 976 AGAAATCCCTTAACCAATGCAATGATCAGGACGACGACGACGACGACGACGACGAC 1035
DB      |||
QY 308 LysSerProAlaGlyAlaAlaMetLeuHisThrThrIleAlaProThrMetLeuLysGly 327
DB      |||
QY 1036 GGGGTCATCAATGATCATCCCGCCAGTGGCCGACGACGACGACGACGACGACGAC 1095
DB      |||
QY 328 SerProLysGluAsnValLeuProGlnAspAlaThrAlaTrpIleAsnTyrGlyIleAla 347
DB      |||
QY 1096 CCTGACAGACAGATCCAGAGGCTCTAGAACCTCAGACGACGACGACGACGACGACGAC 1155
DB      |||
QY 348 ProGlyLysAspSerSerAspLysValMetAlaLysAlaLysGluAlaValGlyLysPro 367
DB      |||
QY 1156 GTCCAGTTCCAGTGTGTGAGTCCCTT-----GACCCCGTCCCGCTCAGC 1200
DB      |||
QY 368 ValGluLeu-----AlaPheGluGlyHisAspGlnLysProSerAlaValSer 383
DB      |||
QY 1201 CCTTCTGATGAC-----AAGGCTTGGGCTACGACGCTCTCCGACGCGCTACAG 1251
DB      |||
QY 384 SerThrThrSerAspAlaTrpLysThrLeuAlaGlyLeuAlaAlaAspLysSerGlnAla 403
DB      |||
QY 1252 TCCGCTGTCCTCCGAGATCAATATTTACTGCCCCAGTTACTTCTTATGGCAGACAGACG 1311
DB      |||
QY 404 ProValValProGlyLeu---ValThrAla-----GlyThrAspSer 416
DB      |||
QY 1312 CGATTCCTTAAACAACTCAGCACTGAGCTCTACAGTTCTACCCCATCTACATACAGCT 1371
DB      |||
QY 417 ArgTyrMetGlyLysLysSerSerAspValTyrArgPheGlnProLeuValLeuThrVal 436
DB      |||
QY 1372 GAAGACTTCAACGACATTCAGATGAGTCAAGAAAATCTCAGTCCAGGCTATGAGACG 1431
DB      |||
QY 437 AspGlyThrLysValIleHisGlyThrAspGlnHisIleSerLeuAspAsnValGluThr 456
DB      |||
```

ID	Accession	Protein Name	Length (AA)	Score	Similarity (%)	Match (%)	DB
QY	1432	CAAGTGAATCATCTTGTAGCTTGATGGAGNATGCT	1467	531.50	40.758	212	QY
Db	457	MetValAlaGlyPheTyrGlnAlaGlyLeuValAlaGluThrAla	468	19.248	137	13	Db
RESULT 6							
ID	013968	PRELIMINARY;	596 AA.				
AC	013968:						
DT	01-JUN-1998	(TREMblrel, 06, Created)					
DT	01-JUN-1998	(TREMblrel, 06, Last sequence update)					
DT	01-JUN-2002	(TREMblrel, 21, Last annotation update)					
DE	Putative carboxypeptidase s precursor (EC 3.4.17.4) (XSCS) (GLY-X carboxypeptidase).						
GN	SPAC249.08						
OS	Schizosaccharomyces pombe (fission yeast).						
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;						
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;						
OC	Schizosaccharomycetes.						
OX	NCBI_TaxID=4896;						
RA	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=972;						
RA	Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;						
RL	Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.						
CC	-1- FUNCTION: NECESSARY FOR USE OF CERTAIN NITROGEN						
CC	SOURCE. MAY ALSO CLEAVE INTRACELLULARLY GENERATED PEPTIDES TO						
CC	-1- CATALYTIC ACTIVITY: PEPTIDYLGLYCINE + H(2)O -> PEPTIDE + GLYCINE.						
CC	-1- PATHWAY: NITROGEN METABOLISM.						
CC	-1- SUBCELLULAR LOCATION: LYSOSOME-LINE VACUOLES (BY SIMILARITY).						
CC	-1- PPM: GLYCOSTATIN (BY SIMILARITY).						
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A; ALSO KNOWN AS THE						
CC	AGE/DAEP/ACYL/CPG2/YSCS FAMILY.						
DR	EMBL: Z98601; CAB11265.1; -						
DR	MEROPS: M20.002; -						
DR	InterPro: IPR002933; Peptidase_M20.						
DR	InterPro: IPR000504; RNA_rec.mot.						
DR	Pfam: PF01546; Peptidase_M20; 1.						
DR	PROSITE: PS00030; RRM_RNP.1; UNKNOWN.1.						
KW	Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;						
FT	Signal.						
FT	CHAIN 1 ?						
FT	CHAIN ?						
FT	CARBOHYD 118 118						
FT	CARBOHYD 466 466						
FT	CARBOHYD 541 541						
FT	CARBOHYD 555 555						
SO	SEQUENCE 596 AA; 66606 MW; D97A3930744364C1 CRC64;						
Alignment Scores:							
Pred. No.:	4.47e-36	Length:	596				
Score:	531.50	Matches:	149				
Percent Similarity:	40.758	Conservative:	91				
Best local Similarity:	25.308	Mismatches:	212				
Query Match:	19.248	Indels:	137				
DB:	3	Gaps:	13				
US-10-014-896-1 (1-1509) x 013968 (1-596)							
QY	28	GCCTGTGGCTATGCTGCTCTTAATTTCCCTACCGTCTCCAGNATGAGGCGCCGAGG	87				
Db	40	AlaAlaValAlaGlyPheLeuLeuLeuLeuLeuLeuLeuValTyrTyrLeuLeuAlaGlyGlySerAla	59				
QY	88	AGCGGAGACATCAAAAGGCGCTCGCGGAMWCCCTTCGTACGTTACACCAAGAGGAACGCGTC	147				
Db	60	AspAsnAspLysGlnGlnLysMetSerHisSerProGlySerCysMetAspSerGluSerAla	79				
QY	148	GGCAGTAAAGAGCGCTGAA-----	168				
Db	80	AlaValSerThrSerAlaLysCysTyrIleProProValLeuThrProAlaLysGluPro	99				
Y	168	-----	168				

Db	100	LysLeuGIyAspAspValSerGIyLeaSerTYrIleArgSerProGIuPhehensnAsp	113
OY	169	-----GGTGCATCCAGATTCCAAAGTGACTTTTAAc-----	201
Db	120	SerLeuValAlaArgPheGlnGluLeuLeuValIleProThrValCysArgAspAspMetGIy	139
OY	202	-----TCGAGAAGTCCAAATACTACAGCCCTGGCGTGGAGTGGAAAAATCATTCATAA	255
Db	140	AspValGIyAspAspAspArgPheAspIlePheIleAlaValAPheGlnAspLysValArgGIu	159
OY	256	GTCTTTCCSTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTGGGAAAGATATAGC	315
Db	160	LeuTYrProAsnIlePheLysLys-----LeuLysValGIuTYrValAsnThrTYrGIy	177
OY	316	CACCGTTTCACATTCACAGAGCTCGAGCCCACTGGACCTTACCTGGTATGGCTCAC	375
Db	178	LeuLeuIleThrLeuGlnGluLeuSerAsnLysAspLeuLysProLeuValLeuMetGIyHis	197
OY	376	TTTGATGTGGTGCCT-----GCCCTGAAGAAGGTGGAGTGGCCCCATTCCT	426
Db	198	GlnAspValValProValAsnGlnAlaSerLeuAspArgTYrTYrPheProProPheSer	217
OY	427	GGTTGGAGCCTGATGGCGTCATCTATAGTTCGGGGCACACTGGACCAAGAAGACTGTGG	486
Db	218	AlaThrTYrHisAsnGIyHisValTYrSerArgGIyAlaIleAspAspLysAsnSerVal	237
OY	487	ATGGGATATACACAGCCCTTGGAGCTCCTGCCTGATCAGAAAGTACATCCCCGGAAGACT	546
Db	238	ValAlaIleLeuGlnAlaLeuGluIleLeuAlaIleSerAspTYrLysProGIuIleThr	257
OY	547	TTCTTCATTTCTCTGGCCATGATGAGGAGTCATCAGAG-----ACAGGGGCTCAGAGATC	603
Db	258	ValIleAlaSerPheGIyPheAspArgGIuGIuValSerGIyTYrArgGIyAlaLeuProLeu	277
OY	604	TCAGCCCTGECTACAGTCAAG-----GGCGTCCAGTACCTTCATTGTGGAC	651
Db	278	AlaHisLysLeuTYrGIuArgTYrGIyLysAspGIyVal-----AlaLeuIleLeuAsp	295
OY	652	GAGGGGGCTTCATCTGGATGATATTCATTCCTTACTCATGAAGCCCATGGCTTGAAT	711
Db	296	GluGIyGIyPheThrIleAsnDeuPheGIyThrLeuPhe-----AlaThrVal	311
OY	712	GCAGTCTCAGAAAGGTTCCATGACCTCATGCTGTCGAAGTAAACATGACTTCAGCCAC	771
Db	312	CysValAlaGIuGIyLysGIyTYrMetAspValHisLeuLysLeuLysThrProGIyHis	331
OY	772	TCTTCAGTCTCCCAAAGAGCAAGCATTTGGCATCTTGCAGCTGTGTCAAGCCGATGG	831
Db	332	AlaSerIleProProProHisThrAsnIleGIyLeuMetSerLysLeuValThrGlnIle	351
OY	832	GAGCAGCACCAATCCTATCATATATTGGAAAGCGGAGACAGTGCATGTATTGGACGA	891
Db	352	GluGIu-----	353
OY	892	CTGGCAATAGAGTTCCCTTCCTTCCTCAATATTAATCTGAGCAACCCATGGCTATTGAA	951
Db	354	-----ProPheGIyGIyLyuLeuThrPheGIuAsnPro-----PheTYr	366
OY	952	CCACTTATAAACAGACTTATGAG-----	975
Db	367	ThrThrLeuGlnCysPheAlaGIuAsnSerAlaAspMetAspAspAsnLeuArgGIuLeu	386
OY	976	-----AGAAATCCCTTTAACCAAT	993
Db	387	IleLysSerGIyAspThrGluLysMetThrAspLeuPheSerLysSerArgLeuTYrArg	406
OY	994	GCAATATATCAGAGCACCAAGCAGCATCCATATATTCAAAGAGGGGTCAAGTTCAATGTC	1053
Db	407	TYrPhePheGIuThrSerIleAlaValAspValIleAsnGIyLysValIyValAsnAla	426
OY	1054	ATCCCCCAGTGGCCCAAGCCACAGTCACTTCGGATTCACCTTGACAGACAGTCCAA	1113

```

Db      427  LeuProGluGluGluThrThrLeuAlaValAlaHisAlaGluValAlaSerLeuGluLeuLys  446
Qy      1114  GAGGCTCGTAGAATCGACAGACAGACATTTGGCTGATATACAGAGCTCCAGTTCGATGTC---  1170
      ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db      447  GluValTyrAspArgTyrGluGluLeuLeuGluGluPheGluHisGluTyrHisValAsn  466
Qy      1170  -----  1170

Db      467  ValThrLeuPheAsnGluGluThrValValGluTyrGluAspAlaIleGluHisIlePhe  486
Qy      1171  -----TTGAGTGCCTTTGACGACCCCTCCGCTCCAGCCCTTGATGACAGGCTTG  1221
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      487  AlaSerThrAlaLysThrLeuGluProSerProValSerProTyrAspGluSerAsp  506
Qy      1222  GGCTACCGAGCTGCTCCCGCCAGACCGGTACAGTCCGCTCCCGGGAAGTCATATATACGCC  1281
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      507  AlaTyrLysLeuAlaGluAlaIleArgTyrThrPheGluAspGluThrSerValThr  526
Qy      1282  CCAGTTCATTCATGGCCACACAGACAGACGCGATTTCTTACAAACCTCCACCATGGCATC  1341
      ||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      527  ProAlaLeuMetProAlaAsnThrAspThrArgHisTyrTrpAsnLeuThrSerAsnIle  546
Qy      1342  TACAGGTTCTACCCCATCTACTACTACAGCTTAAGCTTAAGCTTCAAGCATGCATGGA-----  1399
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      547  TyrArgTrpThrProValSerThrAsnSerThrSerLysAsnSerPheAsnGluHisThr  566
Qy      1396  GTCACAGAGAAATATCATGATCCACGATCCAGATGAGACCCAGGAAGTAATTCATGTCAGTTC  1455
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      567  IleAsnGluAsnMetArgTyrAspAlaHisMetLysSerIleGluPhePheTyrAsnPhe  586
Qy      1456  ATTGACAGATGCTGACACAGACCAAGG  1482
      ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      587  IleLeuValSerAspSerGluGluGlu  595

RESULT 7
QBRNM5 PRELIMINARY; PRT; 469 AA.
ID QBRNM5
AC QBRNM5;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Zn metalloprotein.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_Taxid=446;
RN [1]
RP SEQUENCE FROM N.A.
RA Rankin S., Li Z., Isberg R.R.;
RT "Microphage Induced Genes of Legionella pneumophila: Protection from
RT Reactive Intermediates and Solute Imbalance During Intracellular
RT Growth.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480917; AAM00645.1;-.
SQ SEQUENCE 469 AA; 51844 MM; EBFLE46F0D99CFIB CR64;

Alignment Scores:
Pred. No.: 6.82e-28 Length: 469
Score: 434.00 Matches: 140
Percent Similarity: 45.60% Conservative: 88
Best Local Similarity: 28.00% Mismatches: 208
Query Match: 15.71% Indels: 64
DB: 2 Gaps: 17.

US-10-014-896-1 (1-1509) x QBRNM5 (1-469)
Qy      34  GTGGCTATGCTGCTCTACTGTTTCCCTACCGTCTCCAGATCGATGGGCCGAGAGCGGG  93
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      13  LeuAlaLeuAlaLeuIleValGluGluAlaGluAlaGluTyrThrArgPro---Alaasp  31
Qy      94  GAGCATCAAAAGGCGCTCCGCAATCCCTTCTCAGTTCAGCAAGAAAGAAAGAGCGGTGCATG  15
      ||| ||||| ||| ||| ||||| ::|
Db      32  GluGluGluArgAlaGluArgSerIleTyrLysGluPheIleGlu-----  45

```

OY	154	AAAGAGGGCGGTGAAGAGTGGCATTCACATTCACAAAGTGGCTTTTGGCTGTGAAGTCC	213
Db	46	-----LleGlnSerGlyPheThr-----	Gly 53
OY	214	AATACACAGGCGCTGGCTGAGTTCCGA--AAATACATTCATTAAGTCTTCTACAGATG	270
Db	54	SeThrThrProLeuValGluValAlaValLysTyrLeuLysLys-----	68
OY	271	GTCAGACACAGCTTTATTCACAGATGAAGTGTGGAGAGATATGACCACTGTTCACATTC	330
Db	69	-----AlaGlyPheSerAsnGluAspIlePhe-----Val 78	
OY	331	CAAGGCTGGACGCGC-----ACCTGG 351	
Db	79	GlyGlyAlaSerProGlnLysAlaAsnLeuValValArgTyrArgGlyThrGlyAspLys	98
OY	352	CAGCCCTACCTGCTGATGAGCTCACTTGATGTGTGCTGCTCCCTCGAAGAGGCTGGAG	411
Db	99	LysProLeuLeuLeuLeuAlaHisThrAspValAlaGluLeuLysAlaSerAspTrpSer	118
OY	412	GTGCCCCCATCTCTGGGGTGGAGCGATGATGGCGCTCATATGGTGGGGCAGACATGGAC	471
Db	119	MetAspProPheGlnLeuThrGluLysGluLysTyrPheTyrGlyArgGlyThrLeuAsp	138
OY	472	GACACAAAGCTGTGTGATGGCATCTAGCTACAGGCGCTTGACGCTGCTGATCAGAGAG--	538
Db	139	AspLys-----AlaGlnAlaIleIleTyrIleAlaAsnLeuIleGlnTyrLysGlnGlu	156
OY	529	---TACATCGCCCCGAGATCTTTCTGATCTCTCTGGCGCCATGATGGAGTCAAGCG	585
Db	157	GlyIleLysProLysArgAspIleLeuValAlaLeuThrAlaAspGlnGluLysSer--	175
OY	586	ACAGGGGCTCAGAGAGATCTCAGCGCTGCTACAGTCAAGGGGCG-----GTCCAGTCAAGCC	639
Db	176	SerProTyrAsnGlyIleSerThrProLeuLysAsnHisLysAspLeuIleGluAlaAsp	195
OY	640	TTTCATTTGGAGAGAGGGGGCTTCATCTGTGATATTTCAATCTTAATCTTCAAGAGGCC	699
Db	196	PheAlaLeuAsnGlnGlyLysTyr-----GlyAspLeuAlaAsnGlyLysLysIle	212
OY	700	ATCGCGCTTGATTCAGCTCAGAGAGAGGGTTCATGACATTCAGCTCAGAGTAAACATG	759
Db	213	SerGlnAsnIleGlnValSerGlnLysTyrIleValAsnTyrAsnLeuGluValAlaAsn	232
OY	760	ACTTCAGCGCCACTCTTCAGCTCTCCAAAGAGACAAAGCATTCGCTCTTCGACGCTGT	819
Db	233	LysGlyGlnHisSerSerLeuThrProThrLysAspAlaIleTyrArgLeuAlaGlyAla	252
OY	820	GTCCAGCGGATTTGGAGACAGACAAATGCGCTATTCATATTGGAGGGGACAGTGGTACT	879
Db	253	LeuGlnArgLeuSerLysPheAsnProLeuLysThrAsnAspValThrAlaIleTyr	272
OY	880	GTATTTCAGCAATGGCAATGAGTTT---CCCTTCCCTGTCATATATATCTTGAGCAAC	936
Db	273	PheLysGlnMetAlaIleIleGluThrGlyProLeuLysThrGluMetValGluAlaSer	292
OY	937	CCATGCGTATTTCAGCACTTATTAAGCAAGCTTATGAGAGAGAAATCCCTTAACCAAGCA	996
Db	293	LysGlyLysSerLysGlnAlaMetGlnGluAlaAspSerSerProGlnThrAsnAla	311
OY	997	ATAATCAGACACACACAGCGCTTACCATATTCACAAAGAGGGGTCAAGTTCAATGTCAATC	1056
Db	312	IleLeuArgThrThrCysGlnGlnThrLeuLeuGlnGlyGlyHisAlaMetAsnLeuLeu	331
OY	1057	CCCCCAGTGGCCCCAGCTTACGTCACCTTCGGGATTCACCCCTGGAGACAGATCCCAAGAG	1111
Db	332	ProGlnLeuAsnLysThrIleAsnCysArgValLeuProGluAspSerProGluMet	351
OY	1117	GTCTCAGACACAAAGACATTTGGCTGATTAAGAGATCCAGATTCAGTGTGGTCAATG	1177
Db	352	ValGlnGlnSerLeuLysThrAlaIleAsnAspProGluValThrLeuLysArgIleGly	371

OY 1177 GCCTTACGCCCCCTCCGTCAGCCCTTCTGATGACAAAGCCTTGAGTACACAGCTGCTC 1236
 Db 372 LysLeuSerArgGlyProSerSerProLeuSerProGluIleLeu-----LysThrIle 389
 OY 1237 CGCGACACCGGTACAGCCGCTTCCCGGAAGTCATATTTACTGCCAGTTACTTCTATTT 1296
 Db 390 ThrGlnLeuThrGlnArgGlyTrpProGluValPro---ThrLeuProIleMetValThr 408
 OY 1297 GGCACACACAGACGAGCCGATTTTACAAACCCGACACCATGGCATGGCATGAGTTCTACCCC 1356
 Db 409 GlyAlaThrAspGlyArgGlyLeuArgSerValGlyIleProThrTyrGlyValMetGly 428
 OY 1357 ACTCATATACAGCTCGAAGACTTCAACAGCATCATGAGTACAGAAATCTCAGTC 1416
 Db 429 LeuPheLeuAspArgAspSphPhe---ArgAlaHisGlyArgGspGluArgIleSerVal 447
 OY 1417 CAAGCCTATGACACCAAGTGAATTCATCTTTGAGTTGATTAAGATGCTGACACAGAC 1476
 Db 448 GluSerPheTyrGluAlaHisAlaPheLeuTyrAspLeuValLysGlnLeuSerSerAsn 467
 RESULT 8
 OY2B17 PRELIMINARY; PRJ: 443 AA.
 AC 092B17;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative aminocyclase.
 GN SC06487 OR SC9C7.23.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denapalite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RT Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrett B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)." ;
 RL Nature 417:141-147(2002).
 DR EMBL: AL035161; CAA22735.1; -;
 DR InterPro: IPR001261; ARGE_DAPF_CPG2.
 DR Pfam: PF01546; Peptidase_M20; 1.
 DR PROSITE: PS00758; ARGE_DAPF_CPG2.1; 1.
 SQ SEQUENCE 443 AA: 47786 MW: 072FA33E63F49421 CRC64;

Alignment Scores:
 Pred. No.: 8, 63e-22 Length: 443
 Score: 361.50 Matches: 134
 Percent Similarity: 45.63% Conservative: 75
 Best Local Similarity: 29.26% Mismatches: 181
 Query Match: 13,098 Indels: 68
 DB: 16 Gaps: 20
 US-10-014-896-1 (1-1509) x 092B17 (1-443)
 OY 156 AGAGGCGGTGAAGAAGGCGCATTCAGATTCGCAACAGTACTTTTACGTAAGAGTCCA 215
 Db 34 ArgGlyGlyGlyLeuAspCysGln-GluArgProAlaAlaGluTyrAlaAlaAlaArg----- 51
 OY 216 TACTACAGCCCTGGCTGAGTTGCGAAATTCATTCATTAAGCTTTCTTACAGTGGTCAG 275
 Db 52 -----LeuAlaGluAlaGly-----IleGluProThrLeuLeuG 63
 OY 276 CACACGCTTATTCAGCATGAAAGTGTGGAAGATATGACCACTGTTCACTATTCAGAG 335
 Db 63 uArgThrAlaGlyArgGlyThrAsnValAlaAlaArg-----IleGluG 77
 OY 336 CTCGACCCCGACGCTTGACGCCCTTACCTGATGGCTTGTGATGTTGTTGGTGGCCGCC 395
 Db 77 YThrAspProSerAlaAspAlaLeuLeuValHisGlyHisLeuAspValAlaProAlaG 97
 OY 396 TGAACAGGCTGGAGGTCGCCCTCTGCTGGTTGGAGCGTATGAGCGCTCATATGAG 455
 Db 97 uAlaAlaAspTrpSerValHisProPheSerGlyGluIleArgGspGlyValValTyrG 117
 OY 456 TCGGGGACACATGAGCGACGACAGAACTGTGATGCAATTACTGACGCTTGGAGCTCT 515
 Db 117 YArgGlyAlaValAspMetLysAsn-----MetAspAlaMetIleLeuAl 132
 OY 516 GGTGATCAGAGATGATCATC-----CCCGAAGATCTTCTTCAATTCTCT 560
 Db 132 aValValArgAspTrpAlaArgGlyValArgProArgArgAspValValIleAlaP 152
 OY 561 GGGCCATGATGAGGATCATCA-----GGACAGAGG-----GCTCAGAGAGATCTC 605
 Db 152 eThrAlaAspGluGluAlaSerAlaGluAspGlySerGlyPheLeuAlaAspGlnAla 172
 OY 606 AGCCTGCTACAG-----TCAGGGGCGCTCAGCTGACCTTCAATTGTGGACGAGGGGG 659
 Db 172 aAlaLeuPheGluGlyCysThrGluGlyValSer-----GluSerGlyAl 187
 OY 660 CTTCATCTTGGATGAT-----TTCATTCCTAACTTCAAGAGCCCATGCG 704
 Db 187 aPheThrPheHisAspGlyAlaGlyArgGlnPheTyrPro----- 200
 OY 705 CTGTATTCAGTCCAGAGAGGTTCCATGACCAACCTCATCTGCAAGTAAACATGACTTC 764
 Db 201 -----IleAlaAlaGluGluGlyGlyThrGlyTyrPLeuValLeuThrAlaArgGlyAl 219
 OY 765 AGGCACATCTTCAAGCTCTCCCAAGAGACAGACATGGATGCTTGCAGCTGCTGAC 824
 Db 219 aGlyHisGlySerLysValAsnArgGluAsnAlaIleThrArgLeuAlaAlaLeuThr 239
 OY 825 CCGATTGAGCAGACACCAATGCTTATCATATTGGAAGGGAGCGGTGACTGATTT 884
 Db 239 TArgIleGlyAspHisAlaTrpProLeuArgLeu---ThrProThrValArgAlaAlaLe 258
 OY 885 GCACACATGCGCAAAATGAGTTTCCCTCCCTGTCATATTAATCTGAGCAACCCATGGCT 944
 Db 258 uThrGluIleAlaGlyValTyrGlyIleGluThrAsp-----LeuSerAsp----- 273
 OY 945 ATTGAACACATTATAGCAGGTTTATGAGAGAAATCCCTTACCAATGCAATTAATCAG 1004
 Db 274 -ValAspAlaLeuLeuAspLysLeuGlyGlnAlaGlyLysLeuValGluSerThrValArg 293
 OY 1005 GACACACAGCGACTCACCATTCATTAAGACAGGGGTCAGTTCAATGTCATCCCGCAGT 1064
 Db 1005 -----

Db 449 Serleaspala-----LeuLysPheGlyIleAspValLeu 460

RESULT 11

093H22 PRELIMINARY; PRT; 441 AA.

AC 093H22:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Putative peptidase.

OS Streptomyces avermitilis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

ON NCBI_TaxID=33903;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21477403; PubMed=11572948;

RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,

RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osomoe T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,

RT "Genome sequence of an industrial microorganism Streptomyces

RT avermitilis: Deducing the ability of producing secondary

RT metabolites.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

DR EMBL; AB070954; BAB69369.1; .

DR InterPro; IPR001261; ARGE_DAPE_CPG2.

DR Pfam; PF01546; Peptidase_M20.

DR PROSITE; PS00758; ARGE_DAPE_CPG2.1; UNKNOWN.1.

SO SEQUENCE 441 AA; 47835 MW; 277F24D891926815 CRC64;

Alignment Scores:

Pred. No.: 4e-18 Length: 441

Score: 318.00 Matches: 120

Percent Similarity: 43.08% Conservative: 73

Best local Similarity: 26.79% Mismatches: 159

Query Match: 11.51% Indels: 96

Db: 2 Gaps: 17

US-10-014-896-1 (1-1509) x 093H22 (1-441)

QY 226 CTGGCTGAGTTCGGAAATATCATTAAGTCTTCTCAAGTGCAGACACAGCTTT 285

Db 50 LeuAlaGluValGly----- 54

QY 286 ATCCAGCATGAAGTCTGGAAGATAT-----AGCCACTGTTCAATGCCAA 333

Db 55 LeuGluProGlnIlePheGluSerHisGlnGlyArgAlaSerThrValAlaArgIleGlu 74

QY 334 GGTCGGACCCCGCTGGCAAGCCCTGCTGATGGCTCACTTGTATGTGGTCCGCC 393

Db 75 GlyIuAspProSerArgProAlaLeuIleHisGlnHisThrAspValProAla 94

QY 394 CCTGAAGAAGCGCTGGAGGTCGCCCATCTCTCGGCTGGAGCGTGAAGCGTCATAT 453

Db 95 AsnAlaAspTrpThrHisProPheSerGlyIuIleAlaAspGlyCysValTrp 114

QY 454 GGTCGGGCACTGAGCGACAGAACTGTGTGATGCAATTAAGCGCTGGAGCTC 513

Db 115 GlyArgGlyAlaValAlaSerMetLysAsp-----MetAspAlaMetThrLeu 129

QY 514 CTGCTGATCAGGAAGTACATC-----CCCGAGATCTTTCTTCATTTCT 558

Db 130 AlaValAlaArgAspArgLeuArgThrGlyArgArgProAlaAspIleValIleuAla 149

QY 559 CTGGGCGCATGATGAGGTGATCAGGGACA---GGGGCTGAGAGATCAGCCCTGCTA 615

Db 150 PheIuAlaAspGluGluAlaGlyIleThrGlyIleAluLys-----TyrLeuVal 166

QY 616 CAGTCA-----AGGGGCGTCCAGCTTACCTTCATGTTGAGAGAGGGGCGC 660

Db 167 GlnIuHisProAspLeuPheGluGluGlyValIleThrGluAla---IleGlyGluValGly 185

QY 661 TTCATCTTGATGATATTCATTCCTTAACCTGAAGAAGCCCATGCCCTGATTCAGCTCA 720

Db 186 Phe-----SerPheThrValAsnGluLysLeuArgLeuTyrIleValGluThrAla 202

QY 721 GGAAGGGTTCATGCAAGCTGCGCAAGTAACATGATCGTGCAGCTCTCACT 780

Db 203 GlnLysGlyMetHisTrpMetArgLeuThrValAspGlyThrAlaGlyHisSerMet 222

QY 781 CCTCCAAAGAGACAACATTCGTCATCCCTTGCGAGCTGTGCGAGCGATGGAGACA 840

Db 223 ThrAsnAspAspAsnAlaIleThrGluLeuGlyGluAlaValGlyArgLeuGlyArgHis 242

QY 841 CCAATGCTATGATATTGGAAAGCGGACAGTGCTGATGTTGACCAAGTGGCAAT 900

Db 243 ThrTrpProValArgVal---ThrLysThrValArgSerPheLeuAspGluLeuSerAsp 261

QY 901 GAGTTCCCTGCCCTGCAATTAATCCGACACCAATGCGCTATTGAACCACTTATA 960

Db 262 AlaLeuGlyThrGluLeu-----AspProGluAsnMetAspIleThrLeu 276

QY 961 AGCAGGTTTATGAGAGAAATCCCTTAACCAATGCAATATGACAGCACACGCACTC 1020

Db 277 AlaLysLeuGlyGlyIleAlaLysMetValGlyAlaThrLeuArgAsnSerAlaIlePro 296

QY 1021 ACCATATTAAAGCAGGGGTCAAGTCAATGTCATGCCCGCCAGCGCCACAGTC 1080

Db 297 ThrMetLeuGlyAlaGlyTyrLysValAsnValIleProGlyGlnAlaThrAlaHisVal 316

QY 1081 AACTTCGGATTCACCCGTGACAGACAGTCCAAAGAGCTCTGAATCCAGAG----- 1134

Db 317 AspGlyArgPheLeuProGlnHisGluGluGluPheLeuAlaAspLeuAspTrpGlyLeu 336

QY 1135 -----AACATGTGGCTGATTAACAGATCCAGTTC----- 1164

Db 337 GlyProArgValLysArgGluAspValHisGlyAspLysAlaLeuGluThrAspPheAsp 356

QY 1165 -----CATGTGTGAGTGGCTTGGACCCCTCCCGCTCAGC 1200

Db 357 GlyArgLeuValAspAlaMetGlnSerAlaLeuSerAlaGluAspProIleAla----- 374

QY 1201 CCTTGTGATGACAGAGCGCTTGGGCTACAGCTGCTCCGACAGCTACAGTCCGCTTTC 1260

Db 375 -----LysAlaValProTyrMetLeu----- 381

QY 1261 CCGGAAGTCATTAATTAAGTCCGCCCATCTTACTTATTTGGACACAGACGCGATTTT 1320

Db 382 -----SerGlyGlyThrAspAlaLysSerPhe 390

QY 1321 ACAAACTGACACAGTGCATGTACAGTTCTACCCCATGTACATACAGCTGA---GAC 1377

Db 391 AspAspLeuGlyIleArgCysPheGlyPheAlaProLeuIleuArgProIleuAsp 410

QY 1378 TTCAAA---CGCATTCATGAGTGCAGAGAAATCTGATGCCAAGCTATGAGACCCAA 1434

Db 411 PheAlaGlyMetThrHisGlyIleAspGluArgValProValAspLeuLysPheGly 430

QY 1435 GTGAATTCATCTTGAGTTGATT 1458

Db 431 ValArgValLeuAspArgPheIle 438

RESULT 12

006234 PRELIMINARY; PRT; 448 AA.

AC 006234:

DT 01-JUN-1997 (TREMBLrel. 04, Created)

DT 01-JUN-1997 (TREMBLrel. 04, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Hypothetical 48.1 kDa protein (Aminoacylase-1, putative).

GN DABE2 OR RV2141C OR MTCY270.27 OR MT2199.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

ON NCBI_TaxID=1773;

AC 09AD91;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Putative peptidase.
GN SC01676 OR SC152.18C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_Taxid=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA O'Neil S., Harris D.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabbilowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL590507; CAC36379.1; -;
DR InterPro; IPR001261; ARGE_DAPE_CPG2.
DR InterPro; IPR002933; Peptidase_M20.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01546; Peptidase_M20_1.
DR PROSITE; PS00758; ARGE_DAPE_CPG2.1; UNKNOWN_1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 445 AA; 47879 MW; 95C6E4F7FD56C6 CRC64;

Alignment Scores:

Pred. No.:	1.32e-16	Length:	445
Score:	300.00	Matches:	121
Percent Similarity:	44.29%	Conservative:	69
Best Local Similarity:	28.21%	Mismatches:	181
Query Match:	10.86%	Indels:	58
DB:	16	Gaps:	18

US-10-014-896-1 (1-1509) * 09AD91 (1-445)

QY 226 CTGGCTAGTTCGGAAAATACATTCATTAAGTCTTCTTCTACAGTGTGACACCAGCTTT 285
Db 54 LeuAlaGluValGly----- 58

QY 286 ATCCAGCATTAAGTCGCGAAGATAT-----AGCCACTGTCTACTATGCCAA 333
Db 59 LeuAlaProLysIlePheGluSerHisProGlyArgAlaSerThrValAlaArgIleGlu 78

QY 334 GGCCTGGACCCAGCTTGCAGCCCTTACCTGATGCTGCTGCTGCTGCTGCTGCTGCC 393
Db 334 GGCCTGGACCCAGCTTGCAGCCCTTACCTGATGCTGCTGCTGCTGCTGCTGCC 393

Db 79 GlyIuAspProSerArgProAlaLeuIleHisGlyHisThrAspValValProAla 98
QY 394 CCGTAAGAGCGCTGGAGCGTGGCCATTCCTGCGTTGGAGCGGTGATGCGGTATAT 453
Db 99 AsnAlaAspArgThrHisHisProPheSerGlyIuValAlaAspGlyCysValThr 118
QY 454 GCTCGGGGACACCTGGAGGACAGAACTGTGATGCGATTAATGCGACGCGCTGGAGCTC 513
Db 119 GlyArgGlyAlaValAspMetLysAsp-----MetAspAlaMetThrLeu 133
QY 514 CCGTCGATCAGGAATACATC-----CCCGAAGATCTTCTCATTCCT 558
Db 134 AlaValAlaValAspArgLeuArgSerGlyArgLysProProLysAspIleValIleuAla 153
QY 559 CTGGGCGCATGATGAGAGCTATCATAGGACA---GGGCGCTCAGAGATC----- 603
Db 154 PheIuAlaAspGluAlaGlyIuAlaGlyHisThrGlyIuAlaArgHisLeuValAspHisHis 173
QY 604 TCAGCCCTGCTACAGTCAAGGGCGCTCCAGTACCTTATGTCGAGGAGGGGCTTC 663
Db 174 ProAspLeuPheGlu-----GlyValThrGluAla---IleSerGlyValGlyIlePhe 190
QY 664 ATCTTGATGATTCATTCCTTAATCCTTAAGAGCCATGCGCTGATGTCAGTGCACAG 723
Db 191 -----SerPheThrValAsnGluAlaGlyArgLeuIleGluThrAlaGly 207
QY 724 AAGGTTCCATGACCTCATGCTGCAGTAAGATAGCTTCAAGCGCATTCAGCTTC 783
Db 208 LysGlyIleHisThrMetLysLeuThrValAlaGlyThrAlaGlyHisGlySerMetIle 227
QY 784 CCAAGGAGACAGCATTTGGCATCTGTCAGCTGCTGCGACGCGGATTTGGAGCAGACACA 843
Db 228 HisArgAspAsnAlaIleThrGluLeuSerGluAlaValAlaValLeuGlyArgHisThr 247
QY 844 ATGCCATATATTTTGAAGCGGACAGTGTGACTGTATTCAGCACTGGCAATGAG 903
Db 248 PheProValArgVal---ThrLysThrThrArgAlaPheLeuAspGluLeuGlyAspAla 266
QY 904 TTTCCCTCCCTGTCATATTAATCTGAGCAACCCATGGCTATTTGAACCATTAAGC 963
Db 267 LeuGlyThrAspLeu-----AspProGluAspMetGluGlyThrLeuAla 281
QY 964 AGGTTATGGAGAAATCCCTTAACCATGCAATTAATAGGACGACGCGCATCGC 1023
Db 282 LysLeuGlyGlyIleAlaLysLeuIleGlyAlaThrLeuSerAsnThrAlaAsnProThr 301
QY 1024 AATTTCAAGCAGGGGCTCAATTCATATGTCATCCCCAGTGGCCAGCCACAGCTCAAC 1083
Db 302 GlnLeuGlyAlaGlyIuArgLysValAsnValIleProGlyIuAlaThrAlaHisIleAsp 321
QY 1084 TTCCGGATTCACCCCTGGAGACAGACAGTCCAGAGGCTCCAGAACTCAAGAAATGTGTG 1143
Db 322 GlyArgThrLeuProGlyIuGlu---GluGluPheLeuAlaAspValAspArgThrLeu 340
QY 1144 GGTGATTAAC-----AGATCCAGTTCATGATGTTGAGTTCAGCCCTCCCGCGTC 1197
Db 341 GlyProHisValArgTrpGluAspValHisAlaAsnLysAla-----Val 355
QY 1198 AGCCCTTGTATGATGACAGGCGCTGGCTCAGCTGCTCCGCGACGCTACAGTCCGTC 1257
Db 356 GluThrThrPheAspArgLysAlaLeu---ValAspAlaMetGlnSerAlaLeuValAlaGlu 374
QY 1258 TTCCCGGAAGTCATATTTACTGCCCGCATTTACTTATTTGCGACAGACAGCCGATTC 1317
Db 375 AspProAlaAlaLys---AlaValProTyrMetLeuSerGlyIuThrAspAlaLysSer 393
QY 1318 TTTTACAACCTCAACAGCTGCACTGATGCTTCCATCCCATACATACAGCTGGA--- 1374
Db 394 PheAspGluLeuGlyIleArgGlyPheAlaProLeuLysLeuTyrProGluLeu 413
QY 1375 GACTTCAAA---CGCATTCATGAGAGTCAAGAAATTCATCCCAACCTTATGAGACC 1431
Db 414 AspPheAlaGlyMetThrHisGlyValAspArgLysValProValGluGlyLeuGlnPhe 433

OY 1432 CAGTGAATTCATCTTGATTTGATT 1458
 DB 434 GYValAlrGValleuAspArgPheIle 442

RESULT 14

O93R29 PRELIMINARY; PRT; 442 AA.
 AC O93R29;
 DT 01-DEC-2001 (TREMblrel, 19, Created)
 DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
 DE Purative peptidase.
 GN SC00571 OR SC565.03C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID:1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Collins M.R., Harris D.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE-97000351; PubMed-8843436;
 RX Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinasli H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL; AL591823; CAC39633.1;
 DR Interpro; IPR001261; ARGE_DAPE_CPG2.
 DR Pfam; PF01546; Peptidase_M20; 1.
 DR PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
 DR SEQUENCE 442 AA; 47919 MW; 36762DD638C2CA28 CRC64;

Alignment Scores:

Pred. No.: 1,39e-14 Length: 442
 Score: 276.00 Matches: 116
 Percent Similarity: 40.608 Conservative: 87
 Best Local Similarity: 23.208 Mismatches: 167
 Query Match: 9.99% Indels: 130
 DB: 16 Gaps: 19

US-10-014-896-1 (1-1509) x Q93R29 (1-442)

OY 148 GCGATGAAGAGCGCGTGAAGGTGCGATTCAGATTCCAGATGACTTTAGCTCTGAG 207
 DB 11 AlaSerGIuGIuAlaGIuArgGIuValValAlaSerLycuAlaGIuLeuIleArgPheAsp 30

OY 208 AAGTCCAAATACATACAGCC-----CTGGCT 231
 DB 31 ThrSerAsnProThrSerAspGIuArgAlaCysAlaAspTrpValAlaAlaArgLeuAla 50
 OY 232 GAGTTCGGAATATCATTCATTAAGTCTTTCCTACAGTGGTCAGCAGCAGCTTATTCAG 291
 DB 51 GIuAlaGIu-----IleAla 55
 OY 292 CATGAAGTGTGGAA-----GAGTATACCAACGTTTCATTCAGAGCTCG 339
 DB 56 SerGIuLeuValGIuSerAlaProGIuArgAlaAsnValAlaAlaArgIleProGIuAla 75
 OY 340 GACCCAGCTTCAGCCCTACCTGCTGATGGCTCACTTGTATGGTGGCTGCTGGCCGAA 399
 DB 76 AspTrnSerArgGIuAlaLeuLeuValHisGIuHisLeuAspValValProAlaAspAla 95
 OY 400 GAAGCTGGAGAGTCCCGCATCTCTGGGTGGAGCGCTATGGCGCTATATGTCGG 459
 DB 96 AlaGIuTrpArgValProProPheSerGIuIleGIuAspGIuTyLeuTrpGIuArg 115
 OY 460 GGCACACTGACAGCAGACAGACTCTGTATGGCATTCAGAGCGCTTGAGCTCTGCTG 519
 DB 116 GIuAlaIleAspMetLysAspTrpValAlaValMetLeuAlaTrpAlaArgHisPheAla 135
 OY 520 ATCAGGAATACATCCCGGAGATCTTCTTCATTTCTGTGGCGCATAGAGAGTCA 579
 DB 136 ArgTrnGIuTrpArgProAlaArgGIuIleValLeuAlaPheLeuAlaAspGIuAla 155
 OY 580 TCAGGGACA---GGGGCTCAGAGGATTCAGCCCTGCTACAGTCAAGGGCGCTCAGCTA 636
 DB 156 GIuGIuLysPheGIuAlaHis-----TrpLeuValGIuHisArgProGIuLeuPhe 172
 OY 637 GCGCTTCATT-----GTGACAGAGGGGGG---TTCATCTGGATGATTTC 678
 DB 173 AlaGIuValThrGIuAlaIleGIuGIuGIuLysPheSerTrpAlaLeuAspAspTrp 192
 OY 679 -----ATTCTTAATTCAAGAGCCCATCGCTTATTCAGCACTGGCAATGAG 723
 DB 193 ArgArgLeuTyProIleGIuAsnAlaGIuMetAlaTrpMetGIuLeuThrAla 212
 OY 724 AAGGTTCCATGACCTCATGCTGCAATTAACATGACTTGAAGCCACTTTCAGCTCT 783
 DB 213 AlaGIu-----ArgAlaGIuHisGIuSerSerPro 222
 OY 784 CCAGAGAGACAGATTCGATCCCTTGCAGCTGCTGTCAGCCGATGGAGACACACA 843
 DB 223 AsnAspGIuAsnAlaValThrAspLeuAlaIleSerLeuTrpAlaGIuLeuArgHisThr 242
 OY 844 ATGCTTATCATATTGGAGCGGAGACAGTGTGACTGTATTGACAGCACTGGCAATGAG 903
 DB 243 PhePro-----ValArgLeuIleGIu----- 249
 OY 904 TTTCCCTTCCTGTCATATTAATCTGTGAGCAACCCATGCGTATTGAAACACTT----- 957
 DB 250 -----ProValArgAlaLeuLeuAlaIleAlaAlaArgLeuGIuAlaValAspLeu 266
 OY 958 -----ATTAAGAGGTTT 969
 DB 267 AspLeuAspAlaGIuAspLeuGIuAlaGIuLeuAlaLysLeuGIuHisValAlaAspPhe 286
 OY 970 ATGAGAGAAATCCCTTAACCAATGCAATTAATCAGACACACAGGAGCTACCATATTC 1029
 DB 287 MetGln-----ValValLeuArgAsnSerAlaAsnProTrnMetPhe 300
 OY 1030 AAGAGGCGGTCAGTTCAATGTATCCCGCCCATGGCCCGCCAGGCGACAGTCAACTTCGG 1089
 DB 301 ThrAlaGIuTyArgTrnAsnValIleProGIuArgAlaThrAlaArgValAlaAspGIuArg 320
 OY 1090 ATTCACTGAGCAGACAGTCCAGAGAGCTCTAAGATCAGAAAGACATTTGGCGTGAT 1149
 DB 321 PheLeuProGIuHisGIu-----GIuGIuLeuLeuLys----- 331
 OY 1150 AACAGAGTCCAGTTTCATGTGTGATGAGCTTTCAGCCCTCCCGCTCAGCCCTTTCAT 1209

Db 332 -----ThrIleAspAlaLeuLeuProSerValSer 342
 QY 1210 GACAGCGCTGGGCTACGAGCTGCTCCGACAGAC----- 1245
 Db 343 ArgGluTrpValasnlhisaspIleAlaMetGluTrhSerPheaspolyProLeuValasp 362
 QY 1246 -----GTACAGCTCCGCTTCCTCCGGAAGCAATATTAAGTCCCGCAGTTACT 1290
 Db 363 AlaMetCysAlaIalaValArgAlaGluAspProAspGlyHis---ProValProTyrCys 381
 QY 1291 TCGATTGGCAGACAGCCGATTCCTTACAAACCTCAGCCATCTACAGATTC 1350
 Db 382 AspProGlyCylThrAspAlaYsAlaPheTrhLysLeuAspIleArgCysPheGlyPhe 401
 QY 1351 TACCCCATCTACATACAGCTGAA---GACTTCAAACGATC---CATGGAGTCACAG 1404
 Db 402 LysGlyLeuLysLeuProHisAspLeuAspTyrGlyArgLeuPheHisGlyAlaAspGlu 421
 QY 1405 AAAATCTCAGTCCAGCCCTATGAGACCCAGTGAATTCATCTTGAATTTCAGAAAT 1464
 Db 422 ArgValProLeuGluGlyLeuArgPheGlyValArgValMetTrhArgLeuTrpGlnAsp 441
 RESULT 15
 QYVC09 PRELIMINARY; PRT; 401 AA.
 ID QYVC09; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE CG6733 protein.
 GN CG6733.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Ephygotea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephygotea; Drosophillidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 MD MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bayendale J., Beckett L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöckler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mel M.-H., Ipeyama C.,
 RA Jellali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AF03741; AAF56097.1; -.
 DR FLYbase; FBgn0039052; CG6733.
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; 1.
 SQ SEQUENCE 401 AA; 44925 MW; 9766AE568ED94E CRC64;
 Alignment Scores:
 Pred. No.: 1,06e-12 Length: 401
 Score: 253.50 Matches: 116
 Percent Similarity: 38.77% Conservative: 67
 Best Local Similarity: 24.58% Mismatches: 180
 Query Match: 9.18% Indels: 109
 DB: Gaps: 19
 US-10-014-896-1 (1-1509) x QYVC09 (1-401)
 QY 130 AGCAAGAGGAGGCGTGCGCATGAAGGCGCTGAAGGTCGATCCAGATTCACACA 189
 Db 8 AsnAsnGluGluIleLysIlePheArgGluTyrLeuArg-----IleProThr 23
 QY 190 GTGACTTTTAACTGTGAGAAAGTCAATATACAGCCCTGGTGAATTCGAAATAC--- 245
 Db 24 Val-----HisProAspValAspTyrTrhAlaCysValAlaGluPheLeuLysArgGln 40
 QY 247 -----ATTCATAAAGCTTTTCCCTACAGGTCGACAGCAGCTTT 285
 Db 41 AlaSerSerLeuAsnLeuProValGluValValTyrProAlaValGluTrhLysProVal 60
 QY 286 ATCCAGCAGTGAAGTGTGAGAAAGTATAGCACCCGTTGACATTCAGATTCAGGTCGAGACCC 345
 Db 61 Val-----IleIleLysTrpGluGlySerGlnPro 70
 QY 346 AGCTTGACAGCCCTACCTGTGATGCTCAGTTGATGCTGCTGCTGCTGCTGAGAAAGGC 405
 Db 71 GluLeuSerSerIleValLeuAsnSerTrhAspValValProValPheArgGluLys 90
 QY 406 TGGAGAGTCCCGCCCATTTCTCTGGG---TTGAGCGCTGATGGCGGTCATGATGTCGGGGC 462
 Db 91 TrpTrhHisGluProPheSerLeuAspIleAspGluGluArgIlePheAlaArgGly 110
 QY 463 ACACGTGAGCAGCAAGAACCTGTGATGCTATTCAGAGCCGCTGAGCTGCTGCTGATC 522
 Db 111 ThrGlnAspMetLysSerValGlyTrhGlnTyrLeuGlyAlaIleArgLeuLeuLysAla 130
 QY 523 AGGAGTACATCCCGGAGATCTTCTCATTTCTCTGGGCGATGATGAGAGTCAATCA 582
 Db 131 SerGlyPheLysProLysArgLysLeuLysValTrhPheValProAspGluGluTrhCly 150
 QY 583 GGG---ACAGGGGCTCAGAGATCAGACCCGCTCAAGTCAAGAGGGGCTCCAGCTACCC 639
 Db 151 GlnHisLeuGlyMetAlaGluPheValLysTrhAspTyrTyrLysLysMetLysAlaGly 170
 QY 640 TTCATGTTGGAGAGGGGGGCTTCATCTTGATGATTCATTCATCTCAATCTTAAGAACGCC 699
 Db 171 PheSerLeuAspGluGlyAlaTrhSerGluSerAsp----- 182
 QY 700 ATCCGCTTATGTCAGTCTCAGAGAGAGGGTTCATCAATCAATCAATCAATCAATCAATG 759
 Db 183 ValHisHisLeuPheTyrAlaGluArgLeuArgTrpGlyLeuLysLeuLysValSerGly 202
 QY 760 ACTTCAGGCCACTTTTCAGCTCTCTCCAAAGGAGACAGATTCGATCTTCGACAGCTCT 819
 Db 203 TrhSerGlyHisGlySerLeuLeuLeuProAsnTrhAlaGlyValLysLeuAsnTyrLeu 222
 QY 820 GTGACGAGTATGAGAGACACCAATGCTATCATATTTGGAAGGGGACAGTGTGACT 879
 Db 223 ValAsnLysLeuTrhGlu-----PheArgTrhSerGlnVal----- 234

```
OY 880 GTATTGAGCAACTGGCAAAATGAGTTCCCTCCCTGTCATATATATCTGAGCAACCA 939
Db 235 -----GluasnLeuAlaArgAsp----- 240
OY 940 TGGCTATTGAAACCTTATTAAGCAGGTTTATGAGAGAAATCCCTTAACCAATGCATA 999
Db 241 -----SerLeuSerLysGlyAsp 247
OY 1000 ATCAGACCAACGACGACATCACCATATTCAAAGCAGGGGTCAAGTTCATGTCATCCCC 1059
Db 248 Val--ThrtHrValasnLeuThnGlnLeuSerLysGlyValGlnSerAsnValPro 266
OY 1060 CCAGTGGCCAGGCCACAGTCATCTCCGATTCCACCCCTGACAGACAGTCACAAGAGTC 1119
Db 267 ProLeuPheGlnAlaValPheAspLleArgIle----- 277
OY 1120 CTAGACTCACGAAAGAACTGTGGCTGATTAACAGA----- 1155
Db 278 --AlaIleThrValasnValAlaPheGlnLysGlnIleArgAspTrpCysGlu 296
OY 1156 -----GTCAGTTCATGTGTGAGTGCCTTGGACCCCTCCCC--- 1194
Db 297 AlaGlyGlyGlyIleGlnIleAspPhe-----PheGlnLysGluProTyr 311
OY 1195 GTCAGCCCTCTGTGAT--GACAAAGCCCTGGGCTACCCAGCTGCTCCGCCAGACCGTACAG 1251
Db 312 IleGlyProThrLysLeuAspAsnSerAsnProTyrTrpLeu-----AlaValLys 328
OY 1252 TCCGCTTCCCGAAGCAATATTAAGTCC--CCAGTACTCTATGTCACAAACACAGAC 1308
Db 329 AlaAlaIleAspGlnLeuGlyLeuValHisProIleValCysProGlyAlaThrAsp 348
OY 1309 AGCCGATTCCTTACAAACCTCACCCACTGCATCTACAGTTCACCCCATCTACATPACAG 1368
Db 349 SerArgPheIleArgGlyLysGlyThrProAlaIleGlyPheSerProIleIle----- 366
OY 1369 CTTGAAGACTTCAAAGCATCCATGAGTCACAGAGAAATCTCAGTCCAAAGCCTATGAG 1428
Db 367 --AsnThrThrMetArgIleHisAspHisAspLupheLeuGlnAlaAspValTyrLeu 385
OY 1429 ACCCAAGTGAATATCATCTTGAAGTTCAGAAAT 1464
Db 386 AsnGlyIleAspValTyrLysLysIleIleArgAsn 397
```

Search completed: June 27, 2003, 11:02:25
Job time : 152.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 10:54:48 ; Search time 24 Seconds

(without alignments)
3699.932 Million cell updates/sec

Title: US-10-014-896-1

Perfect score: 2762
Sequence: 1 atggtcagcagcggtgctgtt.....ctcactgtcacaaactgtga 1509

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Xgapop 6.0 , Xgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ n2p.model -DEV-xip
-O/cgn2.1/USPRO.spool/US10014896/runat.27062003.104415.10416/app.query.fasta_1.1671
-DB-issued.patents.AA -OFMT-fastan -SUFFIX-n2p.ra1 -MINMATCH=0.1 -LOOPCTL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=43 -DOCLIN=200 -THR_SCORE-pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT-pio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR-US10014896.@cgn.1.1.28.@unat.27062003.104415.10416 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents.AA:*
1: /cgn2.6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2.6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2.6/prodata/1/1aa/5A.COMB.pep:*
4: /cgn2.6/prodata/1/1aa/5B.COMB.pep:*
5: /cgn2.6/prodata/1/1aa/5A.COMB.pep:*
6: /cgn2.6/prodata/1/1aa/5B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	7.3	418	US-09-134-001C-3552	Sequence 3552, Ap
2	190	6.9	373	US-09-814-951A-2	Sequence 2, App11
3	185.5	6.7	408	US-08-127-278-4	Sequence 4, App11
4	185.5	6.7	408	US-08-555-860-4	Sequence 4, App11
5	185.5	6.7	408	US-09-814-951A-4	Sequence 4, App11
6	168	6.1	431	US-09-134-001C-4338	Sequence 4338, Ap
7	154.5	5.6	446	US-09-134-001C-5125	Sequence 5125, Ap
8	138.5	5.0	383	US-09-575-602-4	Sequence 4, App11
9	118.5	4.3	389	US-09-134-001C-3161	Sequence 3161, Ap
10	116.5	4.2	412	US-09-285-055-2	Sequence 2, App11
11	114.5	4.1	430	US-09-134-001C-4302	Sequence 4302, Ap
12	113	4.1	1284	US-09-343-494-9	Sequence 9, App11

c 13	108	4.0	1253	3	US-08-479-722B-4	Sequence 4, App11
c 14	107.5	4.0	1251	5	PCT-US95-02251-3	Sequence 3, App11
c 15	107.5	4.0	1252	1	US-08-199-780-3	Sequence 3, App11
c 16	107.5	4.0	1252	2	US-08-316-650-3	Sequence 3, App11
c 17	107	3.9	402	4	US-09-319-892-4	Sequence 3, App11
18	105	3.8	642	1	US-08-706-936-3	Sequence 3, App11
19	102	3.7	3969	4	US-08-061-376-5	Sequence 5, App11
20	101	3.7	412	4	US-09-423-439-10	Sequence 10, App11
21	101	3.7	643	4	US-09-423-439-26	Sequence 26, App11
22	101	3.7	642	4	US-09-423-439-16	Sequence 16, App11
23	101	3.7	666	4	US-09-423-439-51	Sequence 51, App11
24	101	3.7	673	4	US-09-423-439-32	Sequence 32, App11
25	100.5	3.6	1224	4	US-09-930-872-4	Sequence 4, App11
26	100	3.6	750	4	US-09-165-239A-4	Sequence 4, App11
27	99.5	3.6	862	1	US-08-325-267A-4	Sequence 4, App11
28	99.5	3.6	894	3	US-08-362-525-22	Sequence 22, App11
29	99.5	3.6	894	3	US-08-971-692-15	Sequence 15, App11
30	99.5	3.6	1481	2	US-08-616-844-40	Sequence 40, App11
31	99.5	3.6	1481	2	US-08-599-634-40	Sequence 40, App11
32	99.5	3.6	1481	3	US-08-944-868A-40	Sequence 40, App11
33	99.5	3.6	1481	3	US-08-944-423A-40	Sequence 40, App11
34	99.5	3.6	1481	3	US-08-944-496-40	Sequence 40, App11
35	99.5	3.6	1537	1	US-08-325-267A-2	Sequence 2, App11
36	99.5	3.6	1610	4	US-09-513-783A-22	Sequence 22, App11
37	99	3.6	1125	4	US-09-513-783A-152	Sequence 152, App11
38	98.5	3.6	446	1	US-08-485-216-2	Sequence 2, App11
39	98.5	3.6	446	2	US-09-003-245-2	Sequence 2, App11
40	98.5	3.6	446	3	US-08-853-552-2	Sequence 2, App11
41	98	3.5	647	4	US-09-423-439-60	Sequence 60, App11
42	97.5	3.5	1804	4	US-09-362-336A-4	Sequence 4, App11
43	96.5	3.5	2627	2	US-08-751-183-3	Sequence 3, App11
44	96.5	3.5	2627	2	US-09-060-836-3	Sequence 3, App11
45	96.5	3.5	2627	4	US-09-184-445-3	Sequence 3, App11

ALIGNMENTS

RESULT 1
US-09-134-001C-3552
; Sequence 3552, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3552
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3552

Alignment Scores:

Pred. No.: 1,48e-11 Length: 418
Score: 201.00 Matches: 106
Percent Similarity: 38.66% Conservative: 73
Best Local Similarity: 22.89% Mismatches: 196
Query Match: 7.28% Indels: 88
DB: 4 Gaps: 19

US-10-014-896-1 (1-1509) x US-09-134-001C-3552 (1-418)

QY 178 CAGATTCACCAAGTGAAGTCTTGAAGTGAAGTCAATACACACCCCTGCTGAGTTC 237
Db 1 TysValIyIleIleThrIysAsnGluIuAspCysGluMetThrValIuSerIuGln 20


```

Db      127 ProGluPheHisAlaLeuArgAlaGlyPheAlaLeuSpGluGlyIleAlaAsp-----144
Oy      694 AAGCCATCGCCGCTGATTGAGTC-----TCAGAGAGAGGTTCCATGAAACCCATGCTG 747
Db      145 ---ProThrAspAlaPheThrValPheTyrSerIleTyrSerProTyrPrtPValAlaVal 163
Oy      748 CAACTAAACATGACTTCAGCCACTCTTCACCTCCCAAGAGAGACATGGATGGCATC 807
Db      164 ThrSerThrGlyArgProGlyHisAlaSerAlaPheMetGluAspThr-----179
Oy      808 CTTCAGACTGCTGTACGCCGATTTGGAGACACCAATGCCCTATCATATTGGAAGCGGG 867
Db      180 -----AlaAlaGluTyrLeuHisLys-----186
Oy      868 ACAGTGGTACTGTATTGCAGCAACTGGCAAAATGATTTCCCTCCCTGCATATATATAC 927
Db      187 ValValAsnSerIleLeuAlaPheArgGluLysGluTyr-----GlnArgLeu 202
Oy      928 CTGAGCAACCCATGGCTATTGTAACCACTTATAGCAGGTTATGAGAGAAATCCCTTA 987
Db      203 GlnSerAsnProHisLeuLysGluGlySerVal-----213
Oy      988 ACCATGCAATATATCAGAGACCACACGCGCACTACCATATTCAAAAGCAGGGGTCAAGTTC 1047
Db      214 -----ThrSerValAsnLeuThrLysLeuGluGluGlyValAlaTyr 227
Oy      1048 AATGTCATCCCCCGCCAGGCCACAGTCAAGTTCGCGATTCACCCGAGACAGACA 1107
Db      228 AsnValIleProAlaThrMetSerAlaSerPheAspPheValAlaProAspValAsp 247
Oy      1108 GTCCAGAGAGGTCTAGAACTCACAGAGAAC-----ATTGGCTGATTAACAGAGTCCAG 1161
Db      248 PheLysAlaPheGluGluGlnLeuGlnSerTyrPcysGlnAlaIleGlyLysValThr 267
Oy      1162 TTCATGTGTGATGGCTTGACCCCTCCCTCCCTGACCCCTTGATGACAAAGCCCTG 1221
Db      268 LeuGluPheAlaGlnLysTyrMetHisProGlnValThrProThrAspSerAspPro 287
Oy      1222 GGTACACAGCTCTCCCGCAGACCCGTACAGTCCGTCTCCCGAAGTCATATTACT-- 1278
Db      288 TrpTrpAlaAlaPheSerArg-----ValGlyLysAspMetAsnLeuThrLeu 303
Oy      1279 GCCCAGACTTACTTCTATTGGCAACACAGACAGCCGATTTTACAAACCTCCACACCTGGC 1338
Db      304 GluProGluIleMetProAlaAlaThrAspAsnArgTyrIleArgAlaValGlyValPro 323
Oy      1339 ATTCACAGTCTTACCCCATTCATACAGCT-----GAAGACTTCAAAAGC 1386
Db      324 AlaLeuGlyPheSerProMetAsnArgThrProValLeuLeuHisAspHisAspGluArg 343
Oy      1387 ATCCAT 1392
Db      344 LeuHis 345

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127, 278
; FILING DATE: 27-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 281136/1992
; FILING DATE: 28-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NEIMARK, Sheridan
; REGISTRATION NUMBER: 20,520
; REFERENCE/DOCKET NUMBER: IMAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
;
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-127-278-4
;
; Alignment Scores:
; Pred. No.: 5,27e-10 Length: 408
; Score: 185.50 Matches: 89
; Percent Similarity: 39.14% Conservative: 57
; Best Local Similarity: 23.86% Mismatches: 158
; Query Match: 6.72% Indels: 69
; DB: 1 Gaps: 13
;
; US-10-014-896-1 (1-1509) x US-08-127-278-4 (1-408)
Oy      319 CTGTTCATATGCCAAGGCTGGAGCCCGACCTGACCCCTGATGCTGATGCTACTT 378
Db      62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLysHisThr 81
Oy      379 GATGTGCTGCTCCCTCCCTGAAGAAAGCTGGGAGTCCCTCATCTCTGGCTTG---GAG 435
Db      82 AspValValProValPheLysGlnHisTyrSerHisAspProPheGluAlaPheLysAsp 101
Oy      436 CGTATGCGCTCATCTATGCTGCGGGGACACCTGAGACAAAGACTGTGATGAGCATTA 495
Db      102 SerGluGlyTyrIleTyrAlaArgGlyAlaGlnAspMetLysGlyValSerIleGlnTyr 121
Oy      496 CTGAGGCGCTTGGAGCTCCTGTGATCAGAGAGTACATCCCGAAGATCTTTCTTCATT 555
Db      122 LeuGluAlaValAlaArgArgLeuLysValGluGlyHisArgPheThrArgThrIleHisMet 141
Oy      556 TCTCTGGGCGCATGATGAGAGGTGATCAGGACAGGGGCTCAGAGATTCAGCCCTGCTA 615
Db      142 ThrPheValProAspGluGlyValGlyGly-----151
Oy      616 CAGTCAAGGGGCGCCAGCTA-----GCTTCATTTGGACAGAGGGGGC 660
Db      152 ---HisGlnIleMetGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170
Oy      661 TTCATCTTGATGATTTTCATCTTCACTTCAGAGAGCCATTCGCTGATTCAGATC--- 717
Db      171 PheAlaLeuAspGluGlyIleAlaAsn-----ProThrAspAlaPheThrValPhe 187
Oy      718 ---TCAGAGAGGTTTCATGAAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTGT 774
Db      188 TyrSerGluTyrSerProTyrPrtPValArgValThrSerThrGlyArgProGlyHisAla 207
Oy      775 TCAGCTCTCCCAAGAGAGACAGACATTTGCGATCTTGACAGCTGCTGTCAGCGGATGGAG 834
Db      208 SerArgPheMetGluAspThr-----AlaAlaGluTyrLeuHis 220

```

```

QY 835 CAGACCAACGATGCTATCATATTTGAGACGGGACAGTGTGATGATTCAGCAACTG 894
   :
   :
   :
Db 221 Lys-----ValValAsnSerIleLeuAlaPheArg 230
QY 895 GCMAATGAGTTTCCTCCCTGTCATATTAATCTGAGCAACCCATGGCTATTGGAACA 954
   :
   :
   :
Db 231 GluLysGluTrp-----GlnArgLeuGlnSerAsnProHisLeuLysGluGly 246
QY 955 CTATTAAGCAGTTTATGAGAGAAATCCCTTAACCAATGCAATTAACAGACACACAG 1014
   :
   :
   :
Db 247 SerVal-----ThSerVal 251
QY 1015 GCATCTACCAATTCMAAGCAGGGGCTCAAGTTCATGTCATCCCGCCAGTGGCCAGGCC 1074
   :
   :
   :
Db 252 AsnLeuThrLysLeuGluGlyValAlaLysAsnValIleProAlaThrMetSerAla 271
QY 1075 ACAGTCACCTCCGAGTTCACCTGGACAGACAGTCCAGAGGCTCCAGAACCTACAGAG 1134
   :
   :
   :
Db 272 SerPheAspPheValAlaProAspValAspPheLysAlaPheGluGluGlnLeuGln 291
QY 1135 AAC-----ATTGTGGCTGATACAGAGTCCAGTTCATGTCAGTGCCTTGACCCC 1188
   :
   :
   :
Db 292 SerTrpCysGlnAlaIleGlyGlyValIleThrLeuGlnPheAlaGlnIleTrpMetHis 311
QY 1189 CTCGCCCTGACCCCTTCTGATGACAGAGCCCTGGGCTACAGCTGCTCCGCCAGACCGTA 1248
   :
   :
   :
Db 312 ProGlnValThrProThrAspAspSerAsnProTrpIleAlaIlePheSerArg----- 329
QY 1249 CAGTCCGCTCCCGGAGAGTCAATATTAATCT-----GCCCGAGTACTTCTATTGGCAACCA 1305
   :
   :
   :
Db 330 -----ValCysLysAspMetAsnLeuThrLeuGlnProGluIleMetProAlaAlaThr 347
QY 1306 GACAGCCGATTCCTTACAAACCTCCACCATGCGCATCTACAGGTTCTACCCCATCTACATA 1365
   :
   :
   :
Db 348 AspAsnArgTyrIleArgAlaValGlyValProAlaLeuGlnIlePheSerProMetAsnArg 367
QY 1366 CAGCCT-----GAAGACTTCAACGATCCAT 1392
   :
   :
   :
Db 368 ThrProValLeuLeuHisAspHisAspLysIleArgLeuHis 380

RESULT 4
US-08-555-860-4
: Sequence 4, Application US/08555860
: Patent No. 5585474
: GENERAL INFORMATION:
: APPLICANT: IMARI, Kanso
: APPLICANT: OHTA, Tsunetaka
: APPLICANT: KURIOTO, Masahi
: TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
: TITLE OF INVENTION: PREPARATION OF SAID PROTEIN
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/555,860
: FILING DATE: 13-NOV-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/127, 278
: FILING DATE: 27-SEP-1993
: APPLICATION NUMBER: JP 281136/1992
: FILING DATE: 28-SEP-1992
: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: NEIMARK, SherIdan
: REGISTRATION NUMBER: 20,520
: REFERENCE/DOCKET NUMBER: IMARI-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 408 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-555-860-4

Alignment Scores:
Pred. No.: 5.27e-10 Length: 408
Score: 185.50 Matches: 89
Percent Similarity: 39.14% Conservative: 57
Best Local Similarity: 23.86% Mismatches: 158
Query Match: 6.72% Indels: 69
DB: 1 Gaps: 13

US-10-014-896-1 (1-1509) x US-08-555-860-4 (1-408)
QY 319 CTGTTCACATATCCAGAGGCTCGGACCCAGCTTGACGCCCTACCTGATGCTACATT 378
   :
   :
   :
Db 62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLeuAsnSerHisThr 81
QY 379 GATGTGTCCTGCCCCCTGAGAGAGGCTGGAGGTGCCCATCTCTGGGTTG---GAG 435
   :
   :
   :
Db 82 AspValValProValPheLysGluHisTrpSerHisAspProPheGluAlaPheLysAsp 101
QY 436 CGTGATGGGCTCATCTATGCTGGGGCACACCTGACAGCAAGAACTGTGTGATGSCATTA 495
   :
   :
   :
Db 102 SerGluGlyTyrIleLysAlaArgLysAlaGlnAspMetLysCysValSerIleLeuIle 121
QY 496 CTCGAGCCCTGAGAGCTCTGCTGATACAGAGTACATCCCGGAAGATCTTCTTCAAT 555
   :
   :
   :
Db 122 LeuGluAlaValAlaArgLeuLysValGluGlyHisArgPheProArgThrIleHisMet 141
QY 556 TCTCTGGCCATGATGAGAGATCATCATCAGGAGGAGGCTCAGAGGATCTCAGCCCTGCTA 615
   :
   :
   :
Db 142 ThrPheValProAspLeuGluValGlyGly----- 151
QY 616 CAGTCAAGGGGGCTCAGCTA-----GCCCTCATGTGAGAGAGGGGGC 660
   :
   :
   :
Db 152 ---HisGlnGlyMetGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170
QY 661 TTCATCTTGATGATTCATTCCTTCACTTCAAGAAAGCCCATCGCTGATTCAGTGC--- 717
   :
   :
   :
Db 171 PheAlaLeuAspGluGlyIleAlaAsn-----ProThrAspAlaPheThrValPhe 187
QY 718 -----TAGAAGAGGGTCCATGAACTCATGCTGCAAGTAACATGACTTCAGGCCACTCT 774
   :
   :
   :
Db 188 TyrSerGluArgSerProTrpIlePheValAlaThrSerThrGlyArgProGlyHisAla 207
QY 775 TCAGCTCTCCCAAGAGAGACAGATGTCATCTTGACAGTCTGCTGTCAGGCCAGTTGAG 834
   :
   :
   :
Db 208 SerArgPheMetGluAspThr-----AlaAlaGluLysLeuHis 220
QY 835 CAGACCAACGATGCTATCATATTTGAGACGGGACAGTGTGATGATTCAGCAACTG 894
   :
   :
   :
Db 221 Lys-----ValValAsnSerIleLeuAlaPheArg 230
QY 895 GCMAATGAGTTTCCTCCCTGTCATATTAATCTGAGCAACCCATGGCTATTGGAACA 954
   :
   :
   :
Db 231 GluLysGluTrp-----GlnArgLeuGlnSerAsnProHisLeuLysGluGly 246
QY 955 CTATTAAGCAGTTTATGAGAGAAATCCCTTAACCAATGCAATTAACAGACACACAG 1014
   :
   :
   :
Db 247 SerVal-----ThSerVal 251

```

QY 1015 GCACTACCATATTAAGAGGAGGTCACATGTCATCCCCAGTGGCCAGGCC 1074
|||
Db 252 AsnLeuThrLysLeuGluGlyLalValAlaValIleProAlaThrMetSerAla 271
QY 1075 ACAGTCAATCTCGGAGTACACCTGGACACACAGTCCACAGAGCTCCAGAACACAGAG 1134
|||
Db 272 SerPheAspPheArgValAlaProAspValAspPheLysAlaPheGluGluLeuGln 291
QY 1135 AAC-----ATTGTGGCTGATTAACAGACAGTCCAGTCCATGTCGTGATGCTTGAACCC 1188
|||
Db 292 SerTrpCysGlnAlaAlaGlyGlyLalValThrLeuGluPheAlaGlnLysTrpMetHis 311
QY 1189 CTCCTCCGTCAGCCCTTGTATGACACAGCCCTGGGCTACCACTGCTCCGACAGCCGTA 1248
|||
Db 312 ProGlnValThrProThrAspSerAsnProTrpTrpAlaAlaPheSerArg----- 329
QY 1249 CAGTCGCTCTCCGGAAGTCATATTACT---GCCCACTTACTTCTATTTGGACACACA 1305
|||
Db 330 -----ValCysLysAspMetAsnLeuThrLeuGluProGluIleMetProAlaAlaThr 347
QY 1306 GACAGCCGATTCCTTACAAACCTCACCATGTCATACAGTTCCTACCCCATCTACATA 1365
|||
Db 348 AspAsnArgTrpIleArgAlaValGlyValProAlaLeuGlyPheSerProMetAsnArg 367
QY 1366 CAGCCT-----GAAGACTCAAAAGCATCCAT 1392
|||
Db 368 ThrProValLeuLeuHisAspHisAspGluArgLeuHis 380

RESULT 5
US-09-814-951A-4
; Sequence 4, Application US/09814951A
; Patent No. 6387661
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; FILE REFERENCE: CLO01179
; CURRENT APPLICATION NUMBER: US/09/814, 951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-951A-4

Alignment Scores:
Pred. No.: 5,27e-10 Length: 408
Score: 185.50 Matches: 89
Percent Similarity: 39.14% Conservative: 57
Best Local Similarity: 23.86% Mismatches: 158
Query Match: 6.72% Indels: 69
DB: 4 Gaps: 13

US-10-014-896-1 (1-1509) x US-09-814-951A-4 (1-408)

QY 319 CTGTCACTATACAGAGCTCGACCCAGCTTGGACGCTCTGATGATGCTACATTT 378
|||
Db 62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuAsnSerHisThr 81
QY 379 GATGTGTGCTCCCTCGAAGAGAGCTGGAGAGTGGCCCATTCCTTGGGTTG---GAG 435
|||
Db 82 AspValValProValPheLysGlnHisTrpSerHisAspProPheGluAlaPheLysAsp 101
QY 436 CGTATGGCGATCATATGATGGGGGACACACTGGACGACAGCAAGACTCTGTATGACATTA 495
|||
Db 102 SerGluGlyTrpIleArgAlaValGlyAlaGlnAspMetLysCysValSerIleGlnTrp 121
QY 496 CTGACAGGCTTGAGACTCTGCTGTATCAGAGATACATCCCGAGAGATCTTTCTTCAAT 555
|||
Db 122 LeuGluAlaValArgArgLeuLysValGluGlyHisArgPheProArgThrIleHisMet 141

QY 556 TCTCTGGCCCATGATGAGAGTCATACAGGAGAGGAGCTCAGAGATCTCAGCCCTGCTA 615
|||
Db 142 ThrPheValProAspGluValGlyGly----- 151
QY 616 CAGTCAAGGAGGGCTCAGCTA-----GCCCTCATTTGGAGAGGGGGGC 660
|||
Db 152 ---HisGlnLysMetCyluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170
QY 661 TTCATCTGGATGATTTCACTTCACTTCACTTCAAGAAAGCCCATGCTTGAATTCAGATC--- 717
|||
Db 171 PheAlaLeuAspGluLylLleAlaAsn-----ProThrAspAlaPheThrValPhe 187
QY 718 ---TCAGAGAGGCTTCCATGAACCTCAGTGCAGAGTAAAGATGACTTCAGGCCATCT 774
|||
Db 188 TyrSerGlnArgSerProTrpTrpValArgValThrSerThrGlyArgProGlyHisAla 207
QY 775 TCAGTCTCCCAAGAGAGACAGCANTGGCATCTTGCAGTCTGCTGACGAGTATGGAG 834
|||
Db 208 SerArgPheMetCyluAspThr-----AlaAlaGluLysLeuHis 220
QY 835 CAGACACCAATGCTTATCATATTGGAGAGCGGACAGTGTGATCTATTTGACAGACATG 894
|||
Db 221 Lys-----ValValAsnSerIleLeuAlaPheArg 230
QY 895 GCAAAAGATTTCCCTCCCTGTCATATATCTGACGACACCCATGCTATTTGACCA 954
|||
Db 231 GluLysGluTrp-----GlnArgLeuGlnSerAsnProHisLysGlyGly 246
QY 955 CTATATAGCAGGTTATGAGAGAAATCCCTTAACCAATGCAATATACAGACACACAG 1014
|||
Db 247 SerVal-----ThrSerVal 251
QY 1015 GCACTACCATATTAAGAGGAGGTCACATGTCATCCCCAGTGGCCAGGCC 1074
|||
Db 252 AsnLeuThrLysLeuGluGlyLalValAlaValIleProAlaThrMetSerAla 271
QY 1075 ACAGTCAATCTCGGAGTACACCTGGACACACAGTCCACAGAGCTCCAGAACACAGAG 1134
|||
Db 272 SerPheAspPheArgValAlaProAspValAspPheLysAlaPheGluGluLeuGln 291
QY 1135 AAC-----ATTGTGGCTGATTAACAGAGTCCAGTCCATGTCGTGATGCTTGAACCC 1188
|||
Db 292 SerTrpCysGlnAlaAlaGlyGlyLalValThrLeuGluPheAlaGlnLysTrpMetHis 311
QY 1189 CTCCTCCGTCAGCCCTTGTATGACACAGCCCTGGGCTACCACTGCTCCGACAGCCGTA 1248
|||
Db 312 ProGlnValThrProThrAspSerAsnProTrpTrpAlaAlaPheSerArg----- 329
QY 1249 CAGTCGCTCTCCGGAAGTCATATTACT---GCCCACTTACTTCTATTTGGACACACA 1305
|||
Db 330 -----ValCysLysAspMetAsnLeuThrLeuGluProGluIleMetProAlaAlaThr 347
QY 1306 GACAGCCGATTCCTTACAAACCTCACCATGTCATACAGTTCCTACCCCATCTACATA 1365
|||
Db 348 AspAsnArgTrpIleArgAlaValGlyValProAlaLeuGlyPheSerProMetAsnArg 367
QY 1366 CAGCCT-----GAAGACTCAAAAGCATCCAT 1392
|||
Db 368 ThrProValLeuLeuHisAspHisAspGluArgLeuHis 380

RESULT 6
US-09-134-001C-4338
; Sequence 4338, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08


```

QY 616 CAGTCAAGGGGCTCCAGCTTACCTTATGTGGACGAGGGGGCTTCATCTTGATGAT 675
Db 173 -----LeuaspAlaaspPhgclYtyrAlaValasp--- 182
QY 676 TTCATTCTTAAGTCAAGAGCCCATCGCTGATTGCGCTGCAGAGAAAGGTTCCATG 735
Db 183 -----AlaserLysaspAlaGlyThrThrValAlleGlyAlaProThrGluVal 198
QY 736 AACCTCATGCTGCAAGTAAACATGACTTCAGGCCCATCTTCAGCTCCCTCCAAAGAGACA 795
Db 199 LysIleTyrThrThrIleLysGlyLysThrAlaHisAlaSerThrProLysLysGlyIle 218
QY 796 AGC---ATTGCAATCTTGCAGCTGCTGCAGCCGATTTGGACAGACACCAATGCCATAC 852
Db 219 SerAlaIleasnIleAlaSerLysAlaIleSerArgMetLys----- 232
QY 853 ATATTGGAAGCGGACAGTGTGACTGATTGACAGCACTGGCAAAATGATTCCTTC 912
Db 232 ----- 232
QY 913 CCTGTCAATATATCTCGAGCAACCAGTGGCTATTTGAACCACTATAAGCAGTTTANG 972
Db 232 ----- 232
QY 973 GAGAGAAATCCCTTACCAATGCAATATCAGAGCCACAGGCACTCAATATTCAAA 1032
Db 233 -----LeuGlyGluValaspAlaLeuThrAlaasnIleGlyLysPheHis 248
QY 1033 GCAGGGGTCAAGTCAATGTCATCCCGCCAGTGGCCAGCCAGCACTCAACTTCGGATT 1092
Db 249 GlyGlySerIleThrIleThrIleIleAlaaspGluValThrLeuGluAlaGluAlaArgSer 268
QY 1093 CACCTGACAGACAGACTC-----CAAGAGTCTCTAGAACTCAGC 1131
Db 269 HisaspArgLysIleasnLysGluValLysHisMetLysGluThrPheGlyThrThr 288
QY 1132 AAGAATCTGTGGCTATTAACAGAGTCAGTCCATGTGTGATGCTTGAACCCCTC 1191
Db 289 AlaaspGluLeuGly---GlyGluAlaGluValLeuValGluLysSerLysProGlyPhe 307
QY 1192 CCGCTCAGCCCTTCATGATGACAAGGCTTGGCTCAGCTGCTCCGCGACAGCCGTACAG 1251
Db 308 GluValSerGluAla---AspLysValThrGluThrAlaIleSerSerAlaLeuAlaLeu 326
QY 1252 TCCGCTCTCCCGGAGTCATATATTACTGCCCGCAGTTACTTCAATTGGCAACAGACAGC 1311
Db 327 GlyLeuLysGlyaspPhrCysIleAlaGly-----GlyLysSeraspGly 341
QY 1312 CGATTCTTACAAACCTCACACACTGCATCTACAGTCTACCCCATCTACATACAGCT 1371
Db 342 AsnIleMetaspGlnTyr-----GlyIle-----ProSerValIleLeuGly 355
QY 1372 GAAGACTTCAAGACGATCCATGAGTCAAGAGAAATGTCAGTCCAGCCATATAGAGAC 1431
Db 356 ValGlyTyrGluaspIleHisThrThrSerGluThrIleAlaIleLysaspMetLysMet 375
QY 1432 CAAGTGAATTCATCTTGTGATGTGATTCAG 1461
Db 376 LeuThrArgGlnIleIleLysIleIleGlu 385

```

```

; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: RECOMBINANT L-N-CARBAMOYLASE
; CURRENT APPLICATION NUMBER: US/09/285,055
; CURRENT FILING DATE: 1999-04-02
; EARLIER APPLICATION NUMBER: DE 198 14 813.5
; EARLIER FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 412
; TYPE: PRN
; ORGANISM: Arthrobacter aureus
US-09-285-055-2

Alignment Scores:
Pred. No.: 0.00456 Length: 412
Score: 116.50 Matches: 71
Percent Similarity: 40.238 Conservative: 69
Best Local Similarity: 20.408 Mismatches: 149
Query Match: 4.228 Indels: 59
DB: Gaps: 16

US-10-014-896-1 (1-1509) x US-09-285-055-2 (1-412)
QY 331 CAGGCTGGAGCCCGACGCTTGACGCTTACCTGCTGATGCTCACTTGTGCTGCT 390
Db 68 GluGlyThrAspProGluLeuProAlaIleAlaValGlySerHisPheaspSerVal--- 86
QY 391 GCCCTGAAGAGCGTGGAGGTCGCCCATTTCTCGGTTGGAGCGATGGCGTATC 450
Db 87 -----ArgaspGly----- 89
QY 451 TATGCTGGGCGACACGTGACGACAGAAACTGTGATGAGCACTGACAGCCCTGGAG 510
Db 90 -----GlyMetPheaspGlyThrAlaGlyValValCysAlaLeuGluAlaAlaArg 105
QY 511 CTCCTGCTGATCAGGAAGTACATCCCGGAAAGATCTTC---TTCATTCTCTGGGCCAT 567
Db 107 ValMetLeuGluaspGlyTyrValaspArgHisProPheGluPheIleAlaIleValGlu 125
QY 568 GATGAGAGG-----TCATCAGAGG---ACAGGCGCTCAGAGAGTCTCAGCCCTGCTA 615
Db 127 GluGluGlyAlaArgPheSerSerIleLeuGlyGlyAlaIleAlaIleGlyLeuVal 145
QY 616 CAGTCAAGGGGCTCAGCTACCTTCACTTGTGACGAGGGGCTTCATCTGATGATGAT 675
Db 147 AlaaspArg-----GluLeuaspSerLeuValaspGluaspGlyValSerValArgGln 164
QY 676 TTCATTCTTACTTC---AAGAAGCCCATTCGCTTGAATGCAATGCTCAGAGAAGGTTCC 732
Db 165 AlaAlaThrAlaPheGlyLeuLysProGlyGluLeuGlnAlaAlaAlaArgSerAlaAla 184
QY 733 ---ATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCCATCTGACCTCCCAAG 789
Db 185 AspLeuAlaGlyAlaPheIleIleLeuHisIleGluGlnGly---ProIleLeuGluGlnGlu 203
QY 790 GAGCAAGCAATTTGGCATCTTGCAGCTGCTGCTGACCCGATTTGAGACAGACACCACTGCT 849
Db 204 GlnIleGluIleGlyValValThrSerIleValGly---ValArgAlaLeuArgValAla 222
QY 850 ATCATATTTTGGAAGCGGACAGTGTGACTGTATTGACAGCACTGGCAATGAGTTTCC 909
Db 223 ValLysGlyArgSerAspHisAlaGlyThrThrPromethIleuArgGlnaspAlaLeu 242
QY 910 TTCCTGTCAATATATATCTCGAGCAACCAGTGGCTATTGAAACCACTTATAAGCAGTTT 969
Db 243 ValProAlaAlaLeuMetValArgGlu-----ValAspArgPhe 255
QY 970 ATGAGAGAAATCCCTTACCAATGCAATATCAGAGCCACAGGCACTGACCATATTC 1009
Db 256 Val-----AsnGluIleAlaaspGlyThrValAlaThrValGlyHisLeuThrValAla 273
QY 1030 AAGCAGGGGTCAAGTCAATGTCATCCCGCAGTGGCCAGGCGCACAGTCACTTCGGG 1089

```

```

Db      274 Proglygly-----GIYASGLINValProglyglyValAspPheThrLeuAspLeuArg 291
OY      1090 ATTCACCCGTGACAGACGTCCAGAGAGTCTCTAGACTCAGCAGAACATTTGGCGAT 1149
Db      292 SerProHisGluGluSerLeuArgValLeuIleAspPheSerValMetValGlyGlu 311
OY      1150 AACAGATCCAGTTCATGTGTGTAGTGCCTTTGAC-----CCCTC 1191
Db      312 ValAlaSerGlnAlaGlyValAlaAlaAspValAspPheAsnLeuSerProVal 331
OY      1192 CCCGTGACCCCTTGTATGACAGAGCCCTGGCTACACAGCTGCTCCGCCAGACCCAGAC 1251
Db      332 GlnLeuAlaProThr-----MetValAspAlaValArg 342
OY      1252 TCCGCTTCCCGAAGCAATATTACT---GCCCAAGTACTCTCTATGGCAACACAGAC 1308
Db      343 GluAlaAlaSerAlaLeuGlnPheThrHisArgAspIleSerSerGlyAlaGlnHisAsp 362
OY      1309 AGCGAATTTTACAAACCTCAC 1332
Db      363 SerMetPheIleAlaGlnValThr 370

RESULT 11
US-09-134-001C-4302
; Sequence 4302, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4302
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4302

Alignment Scores:
Pred. No.: 0.00745 Length: 430
Score: 114.50 Matches: 67
Percent Similarity: 36.45% Conservative: 42
Best Local Similarity: 22.41% Mismatches: 129
Query Match: 4.15% Indels: 62
Gaps: 12

US-10-014-896-1 (1-1509) x US-09-134-001C-4302 (1-430)
OY      325 ACATATCAAGCTCGAGACCCAGCTTCACCTACTGCTGATGCTCACTTTGATG 384
Db      74 ThrLeuysGlnHisAsnProLysAlaProLysLeuIleuAsnGlnHisValAspVal 93
OY      385 GTGCTGCTCCCTGAAAGAGC---TTGGAGAGTGGCCCATTCCTGCTGGAGCGGTGAT 441
Db      94 AlSerValAspAspArgGlnTyrTrpGlnTyrProPheHisLeuThrAsnLysAsp 113
OY      442 GCGGTCACTATGCTGGGCGACACATGACAGACAAAGAACTGTGATGGCATTTACGAC 501
Db      114 GluTrpLeuTyrGlyAlaGlyValSerAspMetLysGlyLysMetSerLeuPheTyr 133
OY      502 GCGTGGAGCTCCTGCTGATCAGAGAACTACATCCCGGAAGATCTTTCTTCATTCTCTG 561
Db      134 ValLeuGlnGlnLeuHisGlnAlaGlyGlnArgProGlnLysPheIleValGlnSer 153
OY      562 GGCATGATGAGAGTCAATCAGGAGGCGCTCAGAGATCTCAGCCCTCAGACGTCA 621

```

```

Db      154 ValValGlyGlnGlnValGlyGlnAlaGlyThrLysArg---AlaCysGlnIleGlyPro 172
OY      622 AGGGGGCTCCAGCTACCTTCATGTGAC-----GAGGG 657
Db      173 LysGly---AspLeuAlaLeuValLeuAspThrSerGluAsnGlnAlaLeuGlyGlnGly 191
OY      658 GCGTCACTCTGGATGATTCATTCCTACTTCAAGAGCCATCGCCTTGATTCAGTC 717
Db      192 GlyValIle-----ThrGlyTrpIleThrVal 200
OY      718 TCAGAGAGGTTCCATGACCT---CATGCTCAAGTAACATGACTTTCAGGCA 770
Db      201 LysSerLysAsnThrLeuHisAspGlyAlaArgSerGlnThrIleHisAlaGlyGly 220
OY      771 CTTCTGAGCTCTCCAAAGAGACAG-----797
Db      221 LeuPheGlyAlaSerAlaIleGlnLysMetThrLysValIleGlnSerLeuAsnGlnLeu 240
OY      798 -----CATTGCATCTTCG-----AGCTGCTGACCGCATGAGACAGACACAT 845
Db      241 GluArgHisTrpGlyValMetLysLysSerProGlyMetProGlyAlaAsnThrIle 260
OY      846 GCGTATCATATTGGACGGGACAGTGTGACTGATTTGACAGCACTGGCAATGACT 905
Db      261 AsnProAlaValIleGlnLysArgHisProAlaPheIleAlaAspGluCysArgLeu 280
OY      906 TCCCTCCCTGCATATATATCTGAGCAACCCATGGCATTTGACCATTTATAGCAG 965
Db      281 TrpIleThrValHisTyrLeuProAsnGlu-----SerTyrLeuSer 294
OY      966 GTTTATGGA-----GAGAAATCCCTTAACCAATGCAATATCAGAC-----1007
Db      295 ValValAsnGluIleGlnLysGlyTyrLeuAsnLysValAlaGlnAlaAspValTrpLeuArg 314
OY      1008 -----CACACGGCACTCAACATTCATCAAGAGGGGTCAAGTT 1046
Db      315 GluAsnProLeuGlnPheGlnTrpGlyThrSerMetIleGlnAspLysGlyGluIle 334
OY      1047 -----CAATGTCATCCCGCCAGTGGCGGACGACAGTCAACTCCGATTC 1094
Db      335 PheProSerPheThrValProThrHisProGlyPheLysGlnLeuGlnGluAla 353

RESULT 12
US-09-343-494-9
; Sequence 9, Application US/09343494
; Patent No. 6413741
; GENERAL INFORMATION:
; APPLICANT: Jeeja, Timothy J.
; APPLICANT: Wickenden, Alan
; TITLE OF INVENTION: ICAGEN, Incorporated
; FILE REFERENCE: 018512-001320US
; CURRENT APPLICATION NUMBER: US/09/343,494
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: US 60/091,469
; EARLIER FILING DATE: 1998-07-01
; EARLIER APPLICATION NUMBER: US 60/116,621
; EARLIER FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1284
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; OTHER INFORMATION: Drosophila Elk (delk; Eag (eag-g0)-like K+
; OTHER INFORMATION: gene) protein
US-09-343-494-9

Alignment Scores:
Pred. No.: 0.021 Length: 1284
Score: 113.00 Matches: 90
Percent Similarity: 34.62% Conservative: 63

```


US-08-479-722B-4

Alignment Scores:

Pred. No.: 0.0659 Length: 1253
 Score: 108.00 Matches: 86
 Percent Similarity: 32.09% Conservative: 34
 Best Local Similarity: 22.99% Mismatches: 143
 Query Match: 4.02% Indels: 112
 DB: 3 Gaps: 21

US-10-014-896-1 (1-1509) x US-08-479-722B-4 (1-1253)

```

QY 1158 GACCTCTGATACACCAATGTTCTTCGATGCTTACAGACCTCTGGACTGTCTGTC 1099
DB 910 AspergillusleuconAspergillusleuconAspergillusleuconAspergillusleucon
QY 1098 AGGCTGAATCCGGAAGTGTGACTGTGGCTGGG---CAGTGGGGGATGACATTGAATT 1042
DB 921 -----GTCGCGCysSerLeuValAlaGlyTyrGlyAspHisCysGluLe 936
QY 1041 GACCCCTGCTTTGAAATATGTGAGTGC----- 1015
DB 937 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg 956
QY 1014 ---CGTGGTGTCTGATTATTCGATTGGTTAAGGATTTCTCTCAATAAAGCTCTTAT 958
DB 957 LeuHisSerGlyGlnGlnHisCys-----GluLeuCysIleProAlaHis 971
QY 957 AGTGTGTTCAAAATAGCCATGGGTTGTCACAGATTATATGACAGGGAAGGAATCATAT 898
DB 972 ArgAspIleAspGlyCysIleLeuPheGlyAlaGlyLeuGlyLysCysVal 991
QY 897 TGCAGTGTGTCATAACAGTACACACTGTCGCCCTTCC---AAATATGATAGCATTTGG 841
DB 992 AsnThrGlnProGlyTyrGlyCysTyrCysLysGlnGlyPheTyrTyrAspGlyAsn--- 1010
QY 840 TGCTGTCACAAATAGCCATGGGTTGTCACAGATTATATGACAGGGAAGGAATCATAT 791
DB 1011 ---LeuLeuGlyCysValAspValAspGlyCysLeuAspGlyLeuAsnCysArgAsnGly 1029
QY 790 CTTTGGAGAGAGTGAAGTGGCTTCATGATGTTACTTGACAGCATAGGTTGATGG 721
DB 1030 ValCysGlnAsnThrArgGlyGlyTyrArg---CysAlaCysThrProProAlaGlyTyr 1048
QY 730 AACCTTCTCTGAGA-----CTGCATCAAGGCGATGGCTTCTTGAATGATGATGA 677
DB 1049 SerProAlaGlnAlaGlnCysLeuIleProGlnArgTyrSerThr----- 1063
QY 676 AATCATCAAGATGAAGACCCCGCTGTCACATGAAGGCTACTGAGCCCTTGACT 617
DB 1064 -----ProGln-----ArgAspValLys 1069
QY 616 GTAGCAGGCTGAGATCCTCTGAGCCCTGTCATGATGATCCTCATCATGAGGAG 557
DB 1070 CysAlaGlyAlaSerGlnGlyArgThrAlaCysValTyrProThrAlaGlyPro---- 1088
QY 556 AATGAAGAAGATCTTCGGGGATGACTCTCTGATCAGACAGGAGCTCCAGGCGCTGA 497
DB 1089 ---AlaLeuThrPheAspAspCys-----CysCysArgGlnProArgLeuGly 1103
QY 496 GTAATGCATCACAGAGATCTTGTGCTGAGTGGCCGACCATATGATGAGCCCATCAC 437
DB 1104 -----ThrGlnCysArgPro---CysProProArgGlyThrGlySerGln 1117
QY 436 GCTCCAAACCCAGAAATGGGGCA-----CCTTCCAGCCCTTCTTCAGGGG 392
DB 1118 CysProThrSerGlnSerGlnSerAsnSerPheThrAspThrSerProLeuLeuGly 1137
QY 391 CAGGCAACCATCAAGTAG----- 371
DB 1138 LysSerProArgAspGlnAspSerSerGlnGlnAspSerAspGlnCysArgCysValSer 1157
QY 370 -----CCATCAGCAGAGTAGGGCTGCAAGCTGGGGTCCGAGCCCTTGATAGTGA 323
  
```

```

DB 1158 GlyArgCysValProArgProGlyAlaValCysGlyCysPro----- 1172
QY 322 ACAGGCTGATACCTTCCACAGCTTCATGCTGATTAAGGTGGTGTGACCACTGAG 263
DB 1173 ---GlyGlyPheGlnLeuAspAlaSerArgAla-ArgCysValAspIleAspGlyCysAr 1191
QY 262 GAAAGACTTATGAATGATTTTCCGAACTCCAGGAGGCTGTAGATTGACTTCTGAG 203
DB 1191 g-----GluLeuAsnGlnArg-----GlyLeuLeuGly 1200
QY 202 AGCTAAAGTCACTGTGGAATCTGGATGACACTTTCAG 163
DB 1200 slySerGlnArgCysValAsnThrSerGlySerPheArg 1213
  
```

RESULT 14

PCT-US95-02251-3
 Sequence 3, Application PC/TUS9502251
 GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
 NUMBER OF INVENTION: CELLS
 NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:
 ADDRESSER: Arnold, White & Durkee

STREET: P.O. Box 4433
 CITY: Houston

STATE: Texas
 COUNTRY: United States of America

ZIP: 77210
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

SOFTWARE: Patent Release #1.0, Version
 SOFTWARE: #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/02251

FILING DATE: CONCURRENTLY HERewith
 CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/316,650

FILING DATE: 30-SEP-1994
 CLASSIFICATION:

APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: DMIC009P--
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:

LENGTH: 1251 amino acids
 TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein

PCT-US95-02251-3

Alignment Scores:

Pred. No.: 0.0739 Length: 1251
 Score: 107.50 Matches: 85
 Percent Similarity: 29.97% Conservative: 31
 Best Local Similarity: 21.96% Mismatches: 133
 Query Match: 4.00% Indels: 139
 DB: 5 Gaps: 20

US-10-014-896-1 (1-1509) x PCT-US95-02251-3 (1-1251)

```

QY 1158 GACTCTGTTATCCAGCCCAATGTTCTGCTGAGTCTTAGACCTCTTGACGTCTGTGCC 1099
    |||||
Db 909 AspSerValLeuAlaThrAsnValThrGln----- 919
QY 1098 AGGTGAATCCGGAATGACTGCTGGCGGC---CACTGGGGGATGACATGAACTT 1042
    |||
Db 920 -----GluCysCysSerLeuGlyAlaGlyTrpGlyAspHisCysGluIle 935
QY 1041 GACCCCTGCTTGAATATGATGAGTGC----- 1015
    |||||
Db 936 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg 955
QY 1014 ---CGTGGTCTTCCTGATTTATGCAATGGTAAAGGATTTCTCTCCATAAACCTGCTAT 958
    |||
Db 956 LeuHisSerGlyGlnIleHisCys-----GluIleCysIleProAlaHis 970
QY 957 AAGTGTTCATAATAGCATGGGTGCTGCAGATATATTTAGACAGGGAAGGAACATCAT 898
    :|:
Db 971 ArgAspIleAspGluCysIleLeuPheGlyAlaGluIleCysGlyGlnGlyLysCysVal 990
QY 897 TGCCAGTCTGCAATACAGTACACACACTGCCCGCTCCCAATAT----- 853
    |||
Db 991 AsnSerGlnProGlyTyrCysLysGlnGlyPheTyrTyrAspGlyAsnLeu 1010
QY 852 -----GATAGGCATTTGGTCTGCTCCATCGGCTGACAGACGCTGCAAGCA- 806
    |||
Db 1011 LeuGluCysValAspValAspGluCysLeu-----AspGluSerAsnLysArgAsn 1027
QY 805 -----TGCCATGCTGCTGCTGCTGAGTGAATGAAATCATCCAAATGAAAGCCCC 776
    |||||
Db 1028 GlyValCysGluAsnThrTyrArgLeuProCysAlaCys---ThrProAlaGluTyr 1046
QY 775 AAGATGGCCTGAAGTATGTTTACTTGACAGCATGAGTTTGAACCTTCTCTGAGA 716
Db 1047 SerProAlaGlnAlaGlnCysLeuIle----- 1055
QY 715 CTGCAATCAAGGCGATGGCTCTGGAAGTTAGATGAAATCATCCAAATGAAAGCCCC 656
    |||||
Db 1056 -----ProGluArgTrpSerThr----- 1061
QY 655 CCTGCTCCAAATGAAAGCTAGCTGACGCCCTTGACTAGTACAGAGGCTGATCCCTCT 596
    |||||
Db 1062 -----ProGln-----ArgAspValLysCysAlaGlyAlaSerGluGlu 1074
QY 595 GAGCCCTGCTCCCTGATGACTCTCTATCATCATGAGCCCAAGAAATGAAAGATCTCGGG 536
    :|:
Db 1075 ArgThrAlaCysValTrpGlyProTrpAlaGlyPro-----AlaLeuThrPheAsp 1091
QY 535 GGATGTACTTCTGATCAGCAGAGAGCTCCAGGCGCTGCAGTATGCCATCAGAGTCT 476
    |||
Db 1092 AspCys-----CysCysArgGlnProArgLeuGly-----ThrGln 1103
QY 475 TGTCTGTCAGTGTGCCCGACCATAGATGACGCCCATCAGCTCCACACCCAGAAATGGGG 416
    |||||
Db 1104 CysArgPro---CysProProArgGlyThrGlySerGlnCysProThrSerGlnSerGlu 1122
QY 415 GCA-----CCTCCAGCCTTCTTCAGGGGCAAGCCACATCAAAAGTGA 371
    :|:
Db 1123 SerAsnSerPheTrpAspThrSerProLeuLeuGlyLysSerProArgAspGluAsp 1142
QY 370 -----CCATCAGCA 362
Db 1143 SerSerGluGluAspSerAspGluCysArgCysValSerGlyProCysValProArgPro 1162
QY 361 GGTAGAGCTGCAAGCTGGGGTCCGAGCCTTGATAGTGAACAGGTGATACTCTTCCA 302
    |||
Db 1163 GlyGlyAlaValCysGluCysPro-----GlyGlyPheGlnLeuAsp 1176
QY 301 GCACTTCATGCTGATTAAGCTGGTGTGACCATCTGAGGAAGACTTATGAAATGATT 242
Db 1177 AlaSerArgAla-ArgCysValAspIleAspGluCysArg----- 1189
QY 241 TTCGAACTCAGCCAGGCGCTGATGATGACTTCTCAGAGCTAAAGTCACTGTTGGA 182

```

```

Db 1190 ----GluLeuAsnGlnArg-----GlyLeuLeuCysLysSerGluArgCysValAs 1205
QY 181 TCTGATGTCGACCTTTCAG 163
Db 1205 nThrSerGlySerPheArg 1211

RESULT 15
US-08-199-780-3
; Sequence 3, Application US/08199780
; Patent No. 5763416
;
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Goldstein, Steven A.
; TITLE OF INVENTION: Gene Transfer Into Bone Cells
; TITLE OF INVENTION: And tissues
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-199-780-3

Alignment Scores:
Pred. No.: 0.0739 Length: 1252
Score: 107.50 Matches: 85
Percent Similarity: 29.97% Conservative: 31
Best Local Similarity: 21.96% Mismatches: 133
Query Match: 4.00% Indels: 139
DB: 1 Gaps: 20

US-10-014-896-1 (1-1509) x US-08-199-780-3 (1-1252)
QY 1158 GACTCTGTTATCCAGCCCAATGTTCTGCTGAGTCTTAGACCTCTTGACGTCTGTGCC 1099
    |||||
Db 910 AspSerValLeuAlaThrAsnValThrGln----- 920
QY 1098 AGGTGAATCCGGAATGACTGCTGGCGGC---CACTGGGGGATGACATGAACTT 1042
    |||
Db 921 -----GluCysCysSerLeuGlyAlaGlyTrpGlyAspHisCysGluIle 936
QY 1041 GACCCCTGCTTGAATATGATGAGTGC----- 1015
    |||||
Db 937 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg 956
QY 1014 ---CGTGGTCTTCCTGATTTATGCAATGGTAAAGGATTTCTCTCCATAAACCTGCTAT 958
    |||
Db 956 LeuHisSerGlyGlnIleHisCys-----GluIleCysIleProAlaHis 971

```

```
OY 957 AAGTGTTCAAATAGCATGGTGTCTCAGATTATATGACAGGAAGGAACATCAT 898
Db 972 ArgaspIleaspIuCsIleleuphegIyAlaIuIleCysIysIuIyCysVal 991
OY 897 TGCAGTTGCTGCATACAGTACACACAGTCCCGCTTCCAAATAT 853
Db 992 AsnserGlnProglYtyrIuCsYtyrCysIysGlnIyPheYtyrAspGlyAsnLeu 1011
OY 852 -----GATAGCATTTGCTCTCTCCATCGGCTGACAGACAGCTGCAAGA - 806
Db 1012 leuGIuCsValaspIaaspGIuCsYleu -----AspIuSerAsnCsYargAsn 1028
OY 805 -----TGCCAAATGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTGAGAGCTG 776
Db 1029 gIyValCysGluAsnThrTrpArgLeuProCysAlaCys ---ThrProProAlaGluTyr 1047
OY 775 AAGATGGCGCTGAAGTATGTTACTTTCAGCATGAGGTTTCATGAACCTCTCTGAGA 716
Db 1048 SerProAlaGlnIaGlnCysLeuIle ----- 1056
OY 715 CTGCATCAAGGCGGTGCTCTTGAAGTTAGGAATGAATCATCCAAATGAGAGCCC 656
Db 1057 -----ProGluArgTrpSerThr ----- 1062
OY 655 CCTGTCCACAAATAGAGCTAGCTGAGGCCCTTGACTGTAGCAGGCTGAGATCCTCT 596
Db 1063 -----ProGln -----ArgAspValIysCysAlaGlyAlaSerGlu 1075
OY 595 GAGCCCTGTCCCTGATGACTCCTCATCAGGCCCGCCAGAAATGAAGAAATCTTGGG 536
Db 1076 ArgThrAlaCysValTrpGlyProTrpAlaGlyPro -----AlaLeuThrPheAsp 1092
OY 535 GGATGTACTTCCATGACAGAGAGCTCAGGCGCTGCAGTAATGCCATCAGAGTCT 476
Db 1093 AspCys -----CysCysArgGlnProArgLeuGly -----ThrGln 1104
OY 475 TGTGTCCAGTGTGCCCGACCATAGATGACGCCCATCAGCTCCACCCAGAGATGGGG 416
Db 1105 CysArgPro ---CysProArgGlyThrGlySerGlnCysProThrSerGlnSerGlu 1123
OY 415 GCA -----CCTCCAGCTTCTTCAAGGGGAGGAGCAGCAGCATCAAGAGAG 371
Db 1124 SerAsnSerPheTrpAspThrSerProLeuLeuGlnIySerProArgAspGluAsp 1143
OY 370 -----CCATCAGCA 362
Db 1144 SerSerGluGluAspSerAspGIuCsYargCysValSerGlyProCysValProArgPro 1163
OY 361 GGTAGGCGCTGCAGCTGGGCTCCGAGCCTTGATAGTGAACAGGTGGCTATCTTCCA 302
Db 1164 GlyGlyAlaValCysGIuCsYPro -----GlyGlyPheGlnLeuAsp 1177
OY 301 CGACTTCATGCTGATAAAGCTGTGCTGACACACTGAGGAAGACTTTATGAATGAT 242
Db 1178 AlaSerArgAla -ArgCysValaspIleaspGIuCsYarg ----- 1190
OY 241 TTCGAACTGACGCGGTGTAGTATTGACTTTCAGAGCTTAAAGTCACTGTTGAA 182
Db 1191 -----GluLeuAsnGlnArg -----GlyLeuLeuCysIysSerGluArgCysValAs 1206
OY 181 TCTGATGAGCAGCTTTCAG 163
Db 1206 nThrSerGlySerPheArg 1212
```

Search completed: June 27, 2003, 11:06:27
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:03:29 ; Search time 69 Seconds

(without alignments)
4795.857 Million cell updates/sec

Title: US-10-014-896-1

Perfect score: 2762
Sequence: 1 atgcgtcagcggctgcgtttg.....ctcactgcacaaactgtga 1509

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Egapop 6.0 , Egapext 7.0
Delep 6.0 , Delext 7.0

Searched: 424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters: 849398

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=nlp
-O=/cgn2_1/USPRO.spool/US10014896/runat_27062003.104417.10539/app.query.fasta.1.1671
-DB=published.Applications_AA -GPM=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human0.cdi -LIST=45 -POCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10014896 -ECN=1.18 -runat_27062003.104417.10539
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-YGAPOP=6 -YGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications_AA*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCY_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCYUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	275.5	10.0	441	9	US-09-738-626-6613
2	190	6.9	773	12	US-10-109-860-2
3	185.5	6.7	408	12	US-10-109-860-4
4	151.5	5.5	476	9	US-10-073-885-103

Adding to
cite

5	151.5	5.5	476	10	US-09-925-301-1397	Sequence 1397, App
6	149.5	5.4	286	9	US-10-073-885-78	Sequence 78, Appl
7	146.5	5.3	403	10	US-09-815-242-10282	Sequence 10282, A
8	145	5.2	449	10	US-09-815-242-5375	Sequence 5375, Ap
9	145	5.2	469	10	US-09-815-242-12290	Sequence 12290, A
10	136.5	4.9	457	9	US-09-738-626-6496	Sequence 6496, Ap
11	136.5	4.9	501	9	US-09-791-389-2	Sequence 2, Appl
12	136.5	4.9	501	9	US-09-791-393-2	Sequence 2, Appl
13	136.5	4.9	501	10	US-09-791-378-674	Sequence 674, App
14	136.5	4.9	507	9	US-10-036-041-57	Sequence 57, App
15	136.5	4.9	507	9	US-10-036-855-57	Sequence 57, App
16	136.5	4.9	507	9	US-09-931-836-57	Sequence 57, Appl
17	136.5	4.9	507	9	US-10-036-214-57	Sequence 57, Appl
18	136.5	4.9	507	9	US-10-036-719-57	Sequence 57, Appl
19	136.5	4.9	507	9	US-10-036-160-57	Sequence 57, Appl
20	136.5	4.9	507	9	US-10-036-958-57	Sequence 57, Appl
21	136.5	4.9	507	9	US-10-036-180-57	Sequence 57, Appl
22	136.5	4.9	507	9	US-10-036-063-57	Sequence 57, Appl
23	136.5	4.9	507	9	US-09-963-290-2	Sequence 2, Appl
24	136.5	4.9	507	12	US-10-036-342-57	Sequence 57, Appl
25	136.5	4.9	508	10	US-09-948-783-139	Sequence 139, App
26	136.5	4.9	508	10	US-09-731-872-242	Sequence 242, App
27	136.5	4.9	509	9	US-09-892-877-137	Sequence 137, App
28	133.5	4.8	471	10	US-09-815-242-10682	Sequence 10682, A
29	132	4.8	466	10	US-09-815-242-13394	Sequence 13394, A
30	132	4.8	466	10	US-09-815-242-13537	Sequence 13537, A
31	123	4.5	190	10	US-09-815-242-5877	Sequence 5877, Ap
32	123	4.5	215	10	US-09-815-242-13061	Sequence 13061, A
33	119	4.3	478	9	US-09-738-626-3883	Sequence 3883, Ap
34	118	4.3	2462	9	US-09-819-104A-5	Sequence 5, Appl
35	114	4.1	890	9	US-09-373-658-4	Sequence 4, Appl
36	113.5	4.1	412	10	US-10-289-360-4	Sequence 6, Appl
37	113.5	4.1	1284	10	US-09-950-772-6	Sequence 9, Appl
38	113	4.1	1284	9	US-10-160-224-9	Sequence 11, Appl
39	113	4.1	1284	9	US-10-188-867-11	Sequence 11, Appl
40	113	4.1	1284	9	US-09-119-855-11	Sequence 11, Appl
41	111	4.0	5877	9	US-10-142-515-11	Sequence 11, Appl
42	110.5	4.0	383	9	US-10-285-074-3	Sequence 3, Appl
43	107	3.9	402	10	US-09-727-238-4	Sequence 4, Appl
44	105	3.8	5935	9	US-10-243-243A-8	Sequence 8, Appl
45	104	3.8	2586	10	US-09-905-129-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-738-626-6613
; Sequence 6613, Application US/09738626
; Publication NO. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAMA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIRO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: 1999-12-16
; PRIOR FILING DATE: 2000-04-07
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6613


```

QY 352 CAGCCCTGACCTGATGCTCACTTGTATGTGGTCCCTCCCTGAGAGAGGCTGGAG 411
    ||||| :||| ||||| ||||| :||| |||
Db 38 SerSerIleuLeuInsSerHisThrAspValProValPheLeuSerHisThrSer 57
QY 412 GTCGCCCATCTCTGGGTG---GAGCGTGAAGGGCTCATCTATGGTGGGGCAACAG 468
    ||||| :||| :||| ||||| |||||
Db 58 HisAspProPheGluAlaPheLeuAspSerGluGlyTyrIleTyrAlaArgValAlaGln 77
QY 469 GACGACAGAACTCTGTGATGGCATTCAGAGCCCTTGAGAGCTGCTGATGACGAG 528
    ||| ||| :||| :||| ||||| ||||| :|||
Db 78 AspMetLeuSerValSerIleGlnTyrLeuGlnAlaValAlaArgIleuLeuValGluGly 97
QY 529 TACATCCCGGAGATCTTCTTCACTTCTGCGGCGCATGATGAGGAGATCATCGAG 588
    :||| :||| :||| :||| :||| :|||
Db 98 HisArgPheProArgThrIleHisMetThrPheValProAspIleuValGluGly--- 116
QY 589 GGGGCTCAGAGATCTCAGCCCTGCTACATCAAGGGGCTCCAGTA----- 636
    :||| :||| :||| :||| :||| :|||
Db 117 -----HisGlnIleuMetGluIleuPheValGlnArg 126
QY 637 ---GCCTTCATGTGGAGAGGGGGCTCATCTGTGATGATTCATTCCTTCACTCAAG 693
    ||| ||| ||||| ||||| :||| |||
Db 127 ProGluPheHisAlaLeuAlaArgAlaGlyPheAlaLeuAspIleuGlyIleAlaAsn----- 144
QY 694 AAGCCCATCCCTGATTCGATC---TCAGAGAGGGTTCATGAACCTCATCTG 747
    ||| ||| :||| :||| :||| :|||
Db 145 ---ProThrAspAlaPheThrValPheTyrSerGlnArgSerProTrpTrpValArgVal 163
QY 748 CAACTAAACATGACTTCAGGCACTGCTTCAGCTCCCTCCAAAGGAGCAACATGCGCATC 807
    :||| :||| :||| :||| :||| :|||
Db 164 ThrSerThrGlyArgProGlnHisAlaSerArgPheMetIleuAspThr----- 179
QY 808 CTTCGACSTGCTGACGCGCATTCGAGCAGACACCAATGCGCTCATATTGTGAGCGG 867
    ||| :||| :||| :||| :||| :|||
Db 180 -----AlaAlaGluLeuSerHisLys----- 186
QY 868 ACAAGTGACTGATTCGACGACACAGTGGCAATGATTCCTCCCTGCAATATATATC 927
    ||| :||| :||| :||| :||| :|||
Db 187 ValValaAsnSerIleuAlaPheArgGluLysGlnTrp-----GlnArgLeu 202
QY 928 CTGAGCAACCCATGATTCATTAAGCACTTAAGCAAGTTTATGAGAGAAATCCCTTA 987
    ||||| ||| ||| :||| :||| :|||
Db 203 GlnSerAsnProHisLysLysGluGlySerVal----- 213
QY 988 ASCAATGCAATATCAGAGCACCAGCCGACATCATTCATTAAGCAGGCGTCAATTC 1047
    ||||| :||| :||| :||| :||| :|||
Db 214 -----ThrSerValAsnLeuThrLysLeuGluGlyValAlaIleTyr 227
QY 1048 AATGATATCCCGGAGTGGCCGACAGTCAATTCGAGATTCACCCCTGGACAGACA 1107
    ||||| ||||| :||| :||| :||| :|||
Db 228 AsnValIleProAlaThrMetSerIleAspPheAspPheArgValAlaProAspValAsp 247
QY 1108 GTCCAAAGGTCCTAGAACTCAGAGAAAC-----ATGTGGGTGATTAACAGAGTCCAG 1161
    :||| :||| :||| :||| :||| :|||
Db 248 PheLysAlaPheGluGlnLeuGlnSerTrpCysGlnAlaAlaGluGlyValThr 267
QY 1162 TTCCATGTGTGAAGCTTGTGACCCGCTCCGCTCCGCTGATTCGACAGGCGCTTG 1221
    :||| :||| :||| :||| :||| :|||
Db 268 LeuIlePheAlaGlnIleArgTrpMetHisProGlnValIleThrProIleArgAspSerAsnPro 287
QY 1222 GCGTACAGCTGCTCCGACAGCCGTAAGTCCGCTCTCCGCGAGTCAATATTAATCT-- 1278
    :||| :||| :||| :||| :||| :|||
Db 288 TrpTrpAlaAlaPheSerArg-----ValLysLysAspMetAsnLeuThrLeu 303
QY 1279 GCGCCAGTACTTATTTGGACAGACAGAGAGCGGCTTTTAAACGCTCCAGCAGGCG 1338
    ||| ||| :||| :||| :||| :|||
Db 304 GluProGluIleMetProAlaAlaThrAspAsnArgTyrIleArgAlaValGlyValPro 323
QY 1339 ATCTACAGAGTTCTGACCCATCATACAGAGCT-----GAAGCTTCAAAAGC 1386
    ||| ||||| :||| :||| :||| :|||
Db 324 AlaLeuGluPheSerProMetAsnArgIleProValLeuLeuHisAspHisAspGlnArg 343
QY 1387 ATCCAT 1392

```

```

Db 344 LeuHis 345
RESULT 3
US-10-109-860-4
; Sequence 4, Application US/10109860
; Patent No. US20020142421A1
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al.
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; FILE REFERENCE: CL001179D1V
; CURRENT APPLICATION NUMBER: US/10/109,860
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/814,951
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-109-860-4

Alignment Scores:
Pred. No.: 5,46e-08 Length: 408
Score: 185.50 Matches: 89
Percent Similarity: 39.14% Conservative: 57
Best Local Similarity: 23.86% Mismatches: 158
Query Match: 6,728 Indels: 69
DB: 12 Gaps: 13

US-10-014-896-1 (1-1509) x US-10-109-860-4 (1-408)
QY 319 CTTGTGATTCGACAGGCTCGGACCCGACCTGACAGCCCTGATGATGCTCATTT 378
    :||| :||| :||| :||| :||| :|||
Db 62 ValLeuThrTrpProGlnTyrAsnProIleuSerSerIleLeuLeuInsSerHisThr 81
QY 379 GATGTGCTGCTCCGCTCCGCTGAGAGGCTGGAGGTGCGCCCATCTCTGGGTG---GAG 435
    ||||| ||||| :||| :||| :||| :|||
Db 82 AspValAlaProValPheLysGlnHisTrpSerHisAspProPheGluAlaPheLysAsp 101
QY 436 CGTGAAGCTCATCTATGTGCGGGGACACTGGACAGCAAAACATCTGTATGCGCATTA 495
    :||| :||| :||| :||| :||| :|||
Db 102 SerGluGlyTyrIleTyrAlaArgGluAlaGlnAspMetLysCysValSerIleGlnTyr 121
QY 496 CTGACGCGCTGAGCTGCTGCGATGACAGAGTCAAGGAGTCCGCGGAGATCTTCTCAT 555
    :||| :||| :||| :||| :||| :|||
Db 122 LeuGlnAlaValAlaArgLeuLeuValGluGlnHisArgPheProArgThrIleHisMet 141
QY 556 TCTGTGGCCATGATGAGAGTCAATCAGGAGGAGGCGTCAAGAGATCTCAGCCCTGTA 615
    :||| :||| :||| :||| :||| :|||
Db 142 ThrPheValProAspIleuValGluGly----- 151
QY 616 CACTCAAGGGGCTCCAGTA-----GCCTTCATTTGTGACAGAGGCGGCG 660
    :||| :||| :||| :||| :||| :|||
Db 152 ---HisGlnIleuMetGluIleuPheValGlnArgProGlnPheHisAlaLeuArgAlaGly 170
QY 661 TTGCATGTGATTCATTCATTCCTACTTCAAGAAACCATGGCGCTGATTCGAGTCC-- 717
    :||| :||| :||| :||| :||| :|||
Db 171 PheAlaLeuAspIleuGlyIleAlaAsn-----ProThrAspAlaPheThrValPhe 187
QY 718 ---TCAGAGAGGGTTCATGAACCTCATGCTCCAAAGTAAAGATGATTCAGGCGCACT 774
    :||| :||| :||| :||| :||| :|||
Db 188 TyrSerGlnArgSerProTrpTrpValArgValThrSerThrGlyArgProGlnHisAla 207
QY 775 TCAAGCTCTCCAAAGAGAGACAGACATTCGAGTCTTGGACSTGCTGACGCGGATGAG 834
    ||| ||| :||| :||| :||| :|||
Db 208 SerArgPheMetIleuAspThr-----AlaAlaGluLysLeuHis 220
QY 835 CAGAGCAACATGCTATCATATTGTGAGAGGAGAGAGTGGATGCTGATTCGACAGAACTG 894
    :||| :||| :||| :||| :||| :|||
Db 221 Lys-----ValValaAsnSerIleuAlaPheArg 230

```


Db 170 ValValSerThrCluProThrAspCysGlnValTyrArgGlyLeuArgGlyLeuArgGlyLeu 189
QY 739 CTCATGCTGCTGAAGTAACATGAGCTTACAGGCACTCTGCTCCATCCACAGAGAGCAAC 798
Db 190 IleArgIleAspValGlnGlyValSerCysHisGlySerIleAlaProGluArgGlyAspAsn 209
QY 799 ATGGACATCTTGACAGCTGCTGTGACGCCATTTGGACAGACACCAATGCTCATATTT 858
Db 210 -----AlaIlePhe 212
QY 859 GGAAGGGGACAGTGGTACTGATTTGCAACAATGCAAAAGTTCCCTCCCTGTC 918
Db 213 LysMetGlyProIleLeuGlyLeuGlnGlyLeuSerGlnArgLeuGlyTyr 230
QY 919 AATATATCTCTGAGCAACCCATGCTATTTGAACACTATAAGAGTTTATGAGAG 978
Db 231 -----AspLeuPheLeuGlyLys 236
QY 979 AATCCTTAACCAATGCAATATATGACAGCACACGACGACCTACCATTTCAAGAGGG 1038
Db 237 Gly-----ThrLeuThrValSerGluIlePhePheThrSer 248
QY 1039 GTCAAGTTCAATGATATCCCCCAAGTGGCCAGGCCACAGTCAACTCCGGATTCACCT 1098
Db 249 ProSerArgCysAlaValAlaAspSerCysAlaValSerIleAspArgArgLeuThrTrp 268
QY 1099 GGACAGACAGTCCACAGAGGCTGAGAACTCAGACAGAAACATTTGCTGATTAACAGATC 1158
Db 269 GlyGluThrTrpLeu----- 273
QY 1159 CAATTCAGTGTGTGAGTGCCTTTGACCCCTCCGCTGACCCCTTGTGATGACAGGCC 1218
Db 274 -----GlyAlaLeuAspGluIleArgAlaLeuProAlaValGlnLysAla 288
QY 1219 TTGGGTACCAAGTGTGCTCGGACAGCCGATGACGCTGCTCCGGAAGTCAATATTA 1278
Db 288 ----- 288
QY 1279 GCCCAGTACTTCTATTGTCACACAGACAGCCGATTTCTTACAAACCTCACACCTGGC 1338
Db 289 AsnAlaValAlaSerMetIleAspArgProSerThrPheGlyLeuValTyrPro 308
QY 1339 ATCTACAGTTTACCCCATCATACATACAGCTGAAGACTCAACAGCATTCAGAGATC 1398
Db 309 ThrGluCysIlePheProThrTrpLysValGlnGluAspHisPheThrVal----- 325
QY 1399 AAGGAGAAATCTCAGTCCAGCTATGAG 1428
Db 326 -----LysAlaLeuValAsnAlaTyrGlu 333

RESULT 8
US-09-815-242-5375
Sequence 5375, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EPI7A.011a
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191, 078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206, 848
PRIORITY FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269, 308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5375
LENGTH: 449
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-5375

Alignment Scores:
Pred. No.: 0.00265 Length: 449
Score: 145.00 Matches: 104
Percent Similarity: 35.00% Conservative: 64
Best Local Similarity: 21.67% Mismatches: 200
Query Match: 5.25% Gaps: 112
DB: 10 Gaps: 19

US-10-014-896-1 (1-1509) x US-09-815-242-5375 (1-449)

QY 163 CTGAAGGTCATCCATGATTCACAGTCTTACCTGAGAACTCCAAATACATA 222
Db 1 LeuLysGlyLeuLeuAlaIleGluSerValArgAspAlaValSerGluAsp-- 19
QY 223 GCCCTGGCGAGTGGGA-----AAATCATTCATTAAGTCTTCTCTACAGTG 270
Db 20 -----AlaProValGlyProGlyProArgLysAlaLeuAspTyrMetGlyLeuAla 37
QY 271 GTACAGCACGCTTATTCACAGATGATGCTGGAAGATATAGCCACTGTCATATC 330
Db 38 HisArgAspGlyPheThrThrHisAspValAspHisIleAlaGlyArgIleGluAlaGly 57
QY 331 CAAGGCTGGAGCCAGCTGACGCCCTACCTGCGATGAGGCTGATGATGATGGCT 390
Db 58 LysGlyAsnAsp-----ValLeuGlyIleLeuCysHisValAspValPro 73
QY 391 GCCCCTGAGAAGGCTGGAGCTGCCCATCTCTGGGTGGAGCGTATGCGCTCATC 450
Db 74 Ala-----GlyAspGlyTrpAspSerAsnProIleGluProValTyrGluAspAlaIle 92
QY 451 TATGTCGGGCGACACTGGACAGACAAAGACTGTGATGCGATTAAGTCAAGCTGGAG 510
Db 93 IleAlaArgGlyThrLeuAspAspLysGlyProThrIleAlaValTyrTyrAlaIleLys 112
QY 511 CTCCTGCTGATCAGGAATGATACCCCGAAGATCTTCTTCATTTCTGCGGCATGAT 570
Db 113 IleLeuGluAspMetAsnValAspTrpLysValGlnIleHisMetIleIleGlyThr 132
QY 571 GAGAGTCTCAGGAGCAGAGGCTCAGAG-----ATCTACGCCCTG 612
Db 133 GluGluSerAspTrpLysCysThrAspArgTyrPheLysThrGluGluMetProThrLeu 152
QY 613 CTACAGTCAAGGCGGCTCAGTACGCTTCAATTTGGACAGAGGGGCTTCACTTGGAT 672
Db 153 GlyPheAlaProAspAlaGluPheProCysIleHisGlyLysGlyIleThrPhe 172
QY 673 GATTCATCTCTAAC----- 687
Db 173 AspLeuValGlnAsnLysLeuThrGluAspGlnAspGluProAspTyrGluLeuIleThr 192
QY 688 TTCAGAAAGCCCATGCGCTTGTATGCACTGACAGTCAAGAGGCTTCATGACCTGCTG 747
Db 193 PheLysSerGlyGluArgTyrAsnMetValProAspHisAlaGluAlaArgValLeuVal 212
QY 748 CAAGTAACATGACT----- 762


```
Db 229 lyscsluasmethrhraspvalleqlnasphneglutyrrhrleucluglnasnhsleu 248
      :::: |||||||
Qy 763 TCAGCCGACCTCTTCAGCTCTCCAAAGAGACAGATTGGCATCTTGCAGCTGCTGC 822
      ||| |||:::
Db 249 Glnclgysperthr-----ValasperclyleuleuValleuthrval 263
      ||| |||:::
Qy 823 AGCCGATTGGACGACAGACCAATCCATATTTGGAAAGCCGGACAGTGTGACTGTA 882
      ||| |||
Db 264 GlnclgylasalaValhlsiglmetaspProserillelvalasnaaglyleuTyrl 283
      ||| |||
Qy 883 TTGAGCACTGGCA-----AATGATTCCTCCCTGCATATA 924
      |||::: ||||| ||| |||
Db 284 leuysrhrleuaslaserleuasnlleuaspasnhsnlaqlalaphne-----Val 300
      ||| |||
Qy 925 ATCTGAGTGAACCCATGGCTATTGAAACCACTTATAGCAGTTTATGAGAGAAATCCC 984
      ||| |||::: ||||| ||| |||
Db 301 AlaPheserhsnrlgtyrleuPheasn-----Serasphegllygu----- 314
      ||| |||
Qy 985 TTAAACAATGCAATATACAGACCCAGCCACTACCAATTCAAAGCGGGCTCAAG 1044
      ||| |||
Db 315 -----AATGTCATCCCCAGTGGCCAGGCCACA-----Lysmetglymetyls 319
      ||| |||
Qy 1045 TTC-----AATGTCATCCCCAGTGGCCAGGCCACA----- 1077
      ||| |||
Db 320 Phehistrhraspvalmetglaspvalthrhrasnillellyalllethrtyraspsn 339
      ||| |||
Qy 1078 -----GTCAACTTCGGATTTCACCCCTGGACAGACAGTCCAAG 1116
      ||| |||
Db 340 GlnasnaaglyleuPhegllylleasnleuargtyrProcluglyrhrleuPheglly 359
      ||| |||
Qy 1117 GTCTGAACTACGAGAAACATTGGCTGATACAGAGTCCAGTTCCTGCTTGAGT 1176
      ||| |||
Db 360 Alametasprrprhealaasnlgullelglntlyrlyrhrleuclgyllys 379
      ||| |||
Qy 1177 GCCTTGACCCCTCCCTCCAGCCCTTGATGACAAAGCCCTGGGCTACAGCTGCTC 1236
      ||| |||
Db 380 ValglrProhrhsrtyrvalasrlyasnasp-----ProPheval 393
      ||| |||
Qy 1237 CGCCAGACCGTACAGTCCGCTTCCCGGAAGTCAAT--ATTACTGCCAGTACTTCT 1293
      ||| |||
Db 394 GlnlysleuvalthratalatyrargasnlglnthrAsnaspmethrgrlProtyrthr 413
      ||| |||
Qy 1294 ATGGCAGACACAGACGCCATCTTACAAACCTCACCCCTGGCATCTACAGTCTTAC 1353
      ||| |||
Db 414 Glyclgylthrtyrlyalaarg-----Asnleuaspysgllyal--Alaphegl 429
      ||| |||
Qy 1354 CCATCTACATACAGCTGAGACTTCAACGCATCATGAGTCAAGAGAAATCTCA 1413
      ||| |||
Db 430 Alametaserhsrsergluasrleu-----MethisglnlysasnlglyrThr 447
      ||| |||
RESULT 10
US-09-738-626-6496
; Sequence 6496, Application US/09738626
; Publication No. US20020197605A1
GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKOTO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6496
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6496
Alignment Scores:
Pred. No.: 0.00157 Length: 457
Score: 136.50 Matches: 80
Percent Similarity: 37.57% Conservative: 50
Best Local Similarity: 23.12% Mismatches: 155
Query Match: 4.94% Indels: 61
          Gaps: 15
US-10-014-896-1 (1-1509) x US-09-738-626-6496 (1-457)
Qy 190 GTGACTTTAGCTTGTGAGAGTCCAAATACTACAGCCCTGCTGAGTGGAAATATACAT 249
      |||::: |||::: |||
Db 29 ValserPheasnSerValhlsrserasrProasnleuclglnasprtyrAlaValalys 48
      ||| |||
Qy 250 CATTAAGCTTTCTTCAAGTGTGCAGACCAACCTTATCCAGATGAAGTCTGGAGAG 309
      ||| |||
Db 49 GlutryVallylglnthrleuthrAsnlaaglyleuthrValsergluPhealaagln 68
      ||| |||
Qy 310 -----TAAAGCCACTGTTCACATATTCAGAGCTGGAGCCAGCTTGACGCCCTAC 360
      ||| |||
Db 69 AsrglythrThrAsnheillellythrargylsrglyserglu--GlyalaProlysal 87
      ||| |||
Qy 361 CTGCTGATGAGTCACTTGTATGTGGCTGCC--CCGAAAGAGCTGGAGGTGCC 417
      ||| |||
Db 88 leuileuThrSerhlsrhraspvalvalProserglProleuAsrleuthrAsprhrsn 107
      ||| |||
Qy 418 CCATCTCTGGGTGGAGCGTGAATGCG-----GTCACTATAGCTGGGGCACA 465
      ||| |||
Db 108 ProPhegluleuthrcluglrgasrAlaaglnhlsgltyrhrgrtyrlyrlyrlyla 127
      ||| |||
Qy 466 CTGGAGACAGAGACGTGTGATGGCATACAGCGCCCTG----- 507
      ||| |||
Db 128 AlaasprlyalysgllyasnleuvalMethlsleuAlaalaLeuargAlaValaLaser 147
      ||| |||
Qy 508 ---GAGCTCTGCTGATCAGAAAGTACATCCCCGAAGTCTTCTTCATTTCTCTGGGC 564
      ||| |||
Db 148 GlyAsprThrThrleuasnleuthrtyrvalAlaGluglyser----- 161
      ||| |||
Qy 565 CATGATGAGGAGTCAATCAGGACAGGGGCTCAGAGAGTCTAGCCCTGATACAGTCAAG 624
      ||| |||
Db 162 -----Glnclgmetgllygllyala-----LeuserAlaileuileuAsprlys 176
      ||| |||
Qy 625 -----GGCTCCAGCTAGCCCTTATCTGGAGAGAGGGGGCTCATCTTGAGT 672
      ||| |||
Db 177 ProgluleuPheaspAlasprAlaileuileuAlaasprSergly----- 191
      ||| |||
Qy 673 GATTTCATCTTAATCTTCAAGAGCCCATCGCTGATGACAGTCAAGAGAGGTTC 732
      ||| |||
Db 192 -----AsnlaservalgllythrProthrleuthrThrThrleuargllyly 207
      ||| |||
Qy 733 ATGAACTCATGCTCAAGTAAACATGACTTCAGGC-----CACTCT----- 774
      ||| |||
Db 208 GlylnValThrValThrValasrPthrleuGluglyAlaValhlsrerglylntlyrly 227
      ||| |||
Qy 775 ---TCAGCTCTCCCAAGAGAGACAGCANTGGCATTCTGACGCTGCTGACCGGATTG 831
      ||| |||
Db 228 GlyAlaAlaPro-----AspAlaValAlaAlaValaArgvalaLeuAsprThrleu 244
      ||| |||
Qy 832 GAGCAGACCAATGCTATCATATTGGAAAGCGGAGAGTGTGACTGTATTCAGCAAA 891
      ||| |||
Db 245 ArgAsprglu-----HsrglyrlythrValilleasprlyalasnhr 256
      ||| |||
```



```

; APPLICANT: Parekh, Rajesh Bhikhu
; TITLE OF INVENTION: Proteins, Genes and their use for
; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)
; FILE REFERENCE: 2543-1-001 N1
; CURRENT APPLICATION NUMBER: US/09/791,393
; CURRENT FILING DATE: 2002-01-02
; EARLIER APPLICATION NUMBER: GB 0004412.3
; EARLIER FILING DATE: 2000-02-24
; EARLIER APPLICATION NUMBER: GB 0030050.9
; EARLIER FILING DATE: 2000-12-08
; EARLIER APPLICATION NUMBER: US 60/254,830
; EARLIER FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(501)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-791-393-2

```

```

Alignment Scores:
Pred. No.: 0.00164      Length: 501
Score: 136.50          Matches: 80
Percent Similarity: 37.30% Conservative: 58
Best Local Similarity: 21.62% Mismatches: 127
Query Match: 4.94%      Indels: 105
DB: 9 Gaps: 17

```

US-10-014-896-1 (1-1509) x US-09-791-393-2 (1-501)

```

QY 334 GGCTCGGACCCGCTGACAGCCCTACCTGATGCTGATGCTGATGCTGCTGCC 393
    |||||
DB 113 GlySerLeuValAspSerSerGlyHisIleLeuVal----- 286
QY 394 CCGTAA---GAAGCTGGAGAGTCCGCCATCTCTGCGTGGAGCTGATGGCGTATC 450
    ::|||
DB 133 AspArgGlyAspGlyTrpLeuThrAspProtyrValLeuThrGluValAspGlyLeu 152
QY 451 TATGCTGGGCGACACTGGAGCAGACAAACTCTGTGATGGCATCTAGCAGGCGCTTGAG 510
    |||||
DB 153 TyrGlyArgGlyAlaThrAspAsnLysGlyProValLeuAlaTrpIleAsnAlaValSer 172
QY 511 CTCCTG---CTGATCAGAGATACATCCCGGAGAGATCTTCTTCATTTCTGCGGCAT 567
    ::|||
DB 173 AlaThrArgAlaLeuGluGlnAspLeuProValAsnIleLysPheIleIleGluGlyMet 192
QY 568 GATGAGAGATCAGTACGAGGAGAGGCGCTCAGAGATCTCAGCCCTAGCTCAAG--- 624
    ::|||
DB 193 GluGlu-----AlaGlySerValAlaLeuGluGluValGluLysGluLys 208
QY 625 -----GGCGTCCAGCTAGCTTCAATTTGTGACAGAGGGGCGCTTCATCTTG 669
    |||||
DB 209 AspArgPhePheSerGlyValAsp---TyrIleVal-----IleSer 221
QY 670 GATGATTTCAATCTTCAATCTCAGAGAGCCGCTGATGATGATGATGATGATGATGATGAT 729
    |||||
DB 222 AspAsnLeuTrpIleSerGlnArgLysPro-----AlaIleThrTyrGlyThrArgGly 239
QY 730 TCCATGACCTCATGCTGCAAGTAAACATGACCTTCAAGC-----CACTCTGAGCTCT 783
    ::|||
DB 240 AsnSerTyrPheMetValGluValLysCysArgAspGlnAspPheHisSerGly----- 287
QY 784 CCAAGAGACAGCATTTGCGCATCTTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843
    |||||
DB 258 -----ThrPheGlyGlyIleLeuHisGluProMetAlaAspLeuValAlaLeuLeu 274
QY 844 ATGCGTATCATATTGAGAGCGGAGACAGTGGTACGTATGATGAGCAAGTGGCAATGAG 903

```

```

DB 275 GlySerLeuValAspSerSerGlyHisIleLeuVal----- 286
QY 904 TTTCCTTCCTTCCTGCAATPATATTCCTGAGCAACCATGCGTATTGCAACCATTAAGC 963
DB 287 -----ProGlyIleThrAspGluValVal----- 294
QY 964 AGTTTATGAGAAATCCCTTAACCAATGCAATATATACAGACACAGGACATCACC 1023
DB 295 -----ProLeuThrGluGluGluHisThr----- 308
QY 1024 ATATTCAAAGCAGGGGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1083
DB 304 ---TyrLysAla----- 306
QY 1084 TTCGGATTCAC-----CTGGACAGACAGTCCAGAGGCTCTTA 1122
DB 307 -----IleHisLeuAspLeuGluGluTyrArgAsnSerSerArgValGluLysPheLeu 324
QY 1123 GAACCTCAGAGAAACATTTGGCGTGAATACAGAGTCCAGTCCATGTTGAGT----- 1176
DB 325 PheAspThrLysGluGluIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis 344
QY 1177 -----GCGTTGACCCCTCCCGCTCAGCCCTTCTGATGATGACAAAGCGCTTGGGC 1224
DB 345 GlyIleGluGlyAlaPheAsp-----GluProGlyThrLysThrValIlePro 360
QY 1225 TACCACTGCTCCCGCAGACCGTACAGTCCGCTTCCCGGAGTCAATATTAAGTCC--- 1281
DB 361 GlyArgValIleGlyLysPheSerIleArgGluValProHisMetAsnValSerAlaVal 380
QY 1282 -----CSAGTTACTTATTTGCAACAGACAGCGCA 1314
DB 381 GluLysGluValThrArgHisLeuGluAspValPheSerLysArgAsnSerSerAsnLys 400
QY 1315 TTCTTACAAACCTCAGCCACTGCAATCTAC 1344
DB 401 MetValValSerMetThrLeuGlyLeuHis 410

```

RESULT 13

```

US-09-791-378-674
; Sequence 674, Application US/09/91378
; Patent No. US20020142303A1
; GENERAL INFORMATION:
; APPLICANT: Parekh, Rajesh
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 9195-061-999
; CURRENT APPLICATION NUMBER: US/09/791,378
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/750,395
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 677
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 674
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (70)..(70)
; OTHER INFORMATION: Xaa = Ile or Leu
US-09-791-378-674

```

```

Alignment Scores:
Pred. No.: 0.00164      Length: 501
Score: 136.50          Matches: 80
Percent Similarity: 37.30% Conservative: 58
Best Local Similarity: 21.62% Mismatches: 127
Query Match: 4.94%      Indels: 105
DB: 10 Gaps: 17

```

US-10-014-896-1 (1-1509) x US-09-791-378-674 (1-501)

PRIOR FILLING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/146970
PRIOR FILLING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILLING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILLING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380142
PRIOR FILLING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/644848
PRIOR FILLING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259
PRIOR FILLING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744
PRIOR FILLING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208
PRIOR FILLING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280
PRIOR FILLING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/874503
PRIOR FILLING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: 09/869599
PRIOR FILLING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/908,827
PRIOR FILLING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US99/10733
PRIOR FILLING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILLING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30720
PRIOR FILLING DATE: 1999-12-22
PRIOR APPLICATION NUMBER: PCT/US00/05601
PRIOR FILLING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILLING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILLING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILLING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILLING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILLING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILLING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: PCT/US00/34956
PRIOR FILLING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILLING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/17800
PRIOR FILLING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: PCT/US01/19692
PRIOR FILLING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: PCT/US01/21066
PRIOR FILLING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: PCT/US01/21735
PRIOR FILLING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 80
SEQ ID NO 57
LENGTH: 507
TYPE: PRT
ORGANISM: Homo Sapien
US-10-036-041-57

Alignment Scores:
Pred. No.: 0.00165
Score: 136.50
Percent Similarity: 37.308
Best Local Similarity: 21.628
Query Match: 4.948
DB: 9
Gaps: 17
US-10-014-896-1 (1-1509) x US-10-036-041-57 (1-507)

QY 334 GCGTGGAGCCCGACCTTGCAGCCCTACCTGATGCTCACTTGTATGTCCTGCC 393
|||||
Db 119 GlySerAspProThrIleuSerGlyThrValCysPheTyrGlyHisIleuAspValGlnProAla 138
QY 394 CCTGAA---GAAGCTGGAGAGTGGCCCATCTCTGCGTTGGAGACCGTGAATGCGGTATC 450
|||
Db 139 AsparGlyAspGlyTyrIleuThrAspProTyrValIleuThrGlnValAspGlyLysIleu 158
QY 451 TATGTCGGGCGACACTGGACGACGAAGAACTGTGATGGCATTAAGTCAAGCCCTGGAG 510
|||||
Db 159 TyrGlyAspGlyAlaThrAspAsnLysGlyProValIleuAlaThrPheIleAsnAlaValSer 178
QY 511 CTCCTG---CTGATCAGAGATGATCCCGGACAGATCTTTCATTTCTCTCGGCCAT 567
|||
Db 179 AlaPheArgAlaIleuGlnIleuAspLeuProValAsnIleuLysPheIleIleGlnIleuTet 198
QY 568 GATGAGAGTCACTCAGAGGAGGCGCTCAGAGAGATCTCAGCCCTCTACATCAAG--- 624
|||||
Db 199 GluIleu-----AlaGlySerValAlaIleuGlnIleuValGlnIleuValGlnIleuLys 214
QY 625 -----GGGTCGACGCTGATCTGATGAGAGGAGGCGCTTCATCTG 669
|||||
Db 215 AsparGlyPhePheSerGlyValAsp-----TyrIleVal-----IleSer 227
QY 670 GATGATTTTCATTTCTTAATCTCAAGAGCCCATCGCTGATTCAGTCTCAGAGAGGCT 729
|||||
Db 228 AsparIleuTyrIleSerGlnArgLysPro-----AlaIleThrTyrGlyThrArgGly 245
QY 730 TCCATGACCTGATGCTGCAATGAATACATGACTTCAAGC-----CACTCTCAGCTCT 783
|||
Db 246 AsnSerTyrPheMetValGlnValLysCysArgAspGlnAspPheHisSerGly----- 263
QY 784 CCAAGAGACAAAGCATTTGCGATCCTTGAGAGCTGGTGCAGGCGATTTGGAGAGACCA 843
|||
Db 264 -----ThrPheGlyGlyIleuHisGlnPheMetAlaAspLeuValAlaIleuLeu 280
QY 844 ATGCCATATCATTTTGGAGCGGACAGTGGTACTGTATTCACCAACTGGCAATGAG 903
|||||
Db 281 GlySerLeuValAspSerSerGlyHisIleuVal----- 292
QY 904 TTTCCTTCCCTGTCATATTAATCCTGAGCAACCCATGGCTATTTGAACCACTTAAGC 963
|||
Db 293 -----ProGlyIleTyrAspGlnValVal--- 300
QY 964 AGGTTATGAGAGAAATCCCTTAACCAATGCATTAAGAGAGACACAGGCACTCACC 1023
|||||
Db 301 -----ProLeuThrGlnIleuGlnIleuAsnThr----- 309
QY 1024 AATTCCAAGCAGGGGTCAAGTTCATGATCCCGGAGTGGCCAGGCACTCAAC 1083
|||||
Db 310 ---TyrLysAla----- 312
QY 1084 TTCGGATTCAC-----CTGGACAGACAGTCCAGAGGCTCTA 1122
|||||
Db 313 -----IleHisLeuAspLeuGlnIleuTyrArgAsnSerSerArgValGlnIleuPheLeu 330
QY 1123 GAAGTCAGAGAAATCTGTGCTGATACAGAGTCCAGTTCATGCTGTGGT----- 1176
|||||
Db 331 PheAspThrLysGlnIleuIleuMetHisLeuTyrPArgTyrProSerIleuSerHis 350
QY 1177 -----GCTTTGACCCCTCCCGCTCAGCCCTTCGATGAGCAAGGCTTGGGC 1224
|||||
Db 351 GlyIleGlnGlyAlaPheAsp-----GlnProGlyThrLysThrValIlePro 366
QY 1225 TACAGACTCTCCGCGACAGCCCTACAGTCCGCTCCGGAAGTAAATTAATCTGCC--- 1281
|||||
Db 367 GlyArgValIleGlyLysPheSerIleArgLeuValProHisMetAsnValSerAlaVal 386
QY 1282 -----CCAGTTACTTCTATGGCAACAGACAGCGGCA 1314
|||||
Db 387 GluLysGlnValThrArgHisIleuGlnAspValPheSerLysArgAsnSerSerAsnLys 406

QY 1315 TTCTTACAACTGCACCTGGCATCTAC 1344
Db 407 MetValValSerMetThrLeuGlyLeuHis 416
RESULT 15
US-10-035-855-57
Sequence 57, Application US/10035855
Publication No. US2003008348A1
GENERAL INFORMATION:
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Stewart, Timothy A.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3030R1C4
CURRENT APPLICATION NUMBER: US/10/035,855
CURRENT FILING DATE: 2001-12-26
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/112514
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114140
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115552
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125826
PRIOR FILING DATE: 1999-03-24
PRIOR APPLICATION NUMBER: 60/127035
PRIOR FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/127706
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/129122
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/130359
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131272
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/13371
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/133379
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132383
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138166
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/144791

PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/146970
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380142
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/644848
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/874503
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: 09/869599
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/908,827
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US99/10733
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30720
PRIOR FILING DATE: 1999-12-22
PRIOR APPLICATION NUMBER: PCT/US00/05601
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: PCT/US00/34956
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/17800
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: PCT/US01/19692
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: PCT/US01/21066
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: PCT/US01/21735
NUMBER OF SEQ ID NOS: 80
SEQ ID NO 57
LENGTH: 507
TYPE: PRT
ORGANISM: Homo Sapien
US-10-035-855-57
Alignment Scores:
Pred. No.: 0.00165
Score: 136.50
Percent Similarity: 37.30%
Best Local Similarity: 21.62%
Query Match: 4.94%
DB: 9
Gaps: 17
Length: 507
Matches: 80
Conservative: 58
Mismatch: 127
Indels: 105
Gaps: 17
US-10-014-896-1 (1-1509) x US-10-035-855-57 (1-507)

QY 334 GCGTCGACCCAGCTTGCAGCCCTTACCTGATGCTCATCTTGATGTGGTCC 393
Db |||||
119 GlySerAspProThrLysGlyThrValCysPheTyrGlyHisLeuAspValGlnProIla 138
QY 394 CCTGAA---GAGGCTGGAGGAGTGGCCCATCTCTGGGTTGGAGGATGAGCCGATC 450
Db |||||
139 AspArgGlyAspGlyTrpLeuThrAspProTyrValLeuThrGlnValAspGlyLeu 158
QY 451 TATGTCGGGGGACACTGAGCAGACAGACTGTGATGGCATTTACTGAGCCTTGAG 510
Db |||||
159 TyrGlyArgGlyAlaThrAspAsnLysGlyProValLeuAlaTrpIleAsnAlaValSer 178
QY 511 CTCCTG---CTGATCAGGAGTACATCCCGGAAAGATCTTTCTCATTTCTCGGGCAT 567
Db |||||
179 AlaPheArgAlaLeuGlnAspLeuProValAsnLysPheIleIleGlnGlyMet 198
QY 568 GATGAGGAGTCAATCAGGAGGAGGCTCAGAGAGTCAAGCCCTGCTACAGTCAMG--- 624
Db |||||
199 GlnGlu-----AlaGlySerValAlaLeuGlnGluLeuValGlnLysGlnLys 214
QY 625 -----GGCGTCCAGCTACGCTTCAATCTGTGACGAGGGGGGCTCATCTTG 669
Db |||||
215 AspArgPhePheSerGlyValAsp-----TyrIleVal-----IleSer 227
QY 670 GATGATTTCAATCTCTTAATCTCAGAGAGCCCATGCGCTTGATTCAGTCTCAGAGAGGT 729
Db |||||
228 AspAsnLeuTrpIleSerGlnArgLysPro-----AlaIleThrTyrGlyThrArgGly 245
QY 730 TCCATGAACTCATGCTGAGCAAGTAAACATACATTCAGGC-----CACTTTCAGCTCCT 783
Db |||||
246 AsnSerTyrPheMetValGluValLysCysArgAspGlnAspPheHisSerGly----- 263
QY 784 CCAAGGAGACAGACATTTGATCCCTTGCAGCTGCTGTCTGACCCGATTTGGAGCAGACACA 843
Db |||||
264 -----ThrPheGlnGlyIleLeuHisGlnProMetAlaAspLeuValAlaLeuLeu 280
QY 844 ATGCTATCATATTTGAGAGCGGGAGCAGTGTGACTGATTTGCAAGCACTGGCAATGAG 903
Db |||||
281 GlySerLeuValAspSerSerGlyHisIleLeuVal----- 292
QY 904 TTTCCTTCCTGTCATATATATATCTGAGCAACCCATGCTATTTGAAACCTATTAAGC 963
Db |||||
293 -----ProGlyIleTyrAspGlnValVal----- 300
QY 964 AGCTTTATGAGAGAAATCCCTTAACCAATGCAATTAAGCAAGCAACGACATGAC 1023
Db |||||
301 -----ProLeuThrGlnGluGlnIleAsnThr----- 309
QY 1024 ATATTCAAGCAGGGGTCAAGTCAATGTCAATCCCCCAGTGGGCCAGCCACAGTCAAC 1083
Db |||||
310 ---TyrLysAla----- 312
QY 1084 TTCCGGATTAC-----CCTGACAGACAGTCCAGAGGTCSTA 1122
Db |||||
313 -----IleHisLeuAspLeuGlnGluTyrArgAsnSerSerArgValGlnLysPheLeu 330
QY 1123 GAACTCAGAGAACATTTGGCTGATGATCAGAGTCCAGATTCATGTGTGAGT----- 1176
Db |||||
331 PheAspTrpLysGlnGluIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis 350
QY 1177 -----GCCTTGACCCCTCCCGTCAAGCCCTTGTGATGACAAAGGCTTGGGC 1224
Db |||||
351 GlyIleGlnGluAlaPheAsp-----GlnProGlyThrLysThrValIlePro 366
QY 1225 TACACAGCTGCTCCCGCAGCCGCTACAGTCCGCTCCCGGAGATCATATTACTGCC--- 1281
Db |||||
367 GlnArgValIleGlnLysPheSerIleArgLeuValProHisMetAsnValSerAlaVal 386
QY 1282 -----CAGTTACTTCTATTTGGCAACACAGACAGCCGA 1314
Db |||||
387 GlnLysGlnValThrArgHisLeuGlnAspValPheSerLysArgAsnSerSerAsnLys 406

QY 1315 TTCTTTACAAACCTCACCACTGGCATCTAC 1344
Db |||||
407 MetValValSerMetThrLeuGlnLeuHis 416

Search completed: June 27, 2003, 11:26:24
Job time : 92 secs

